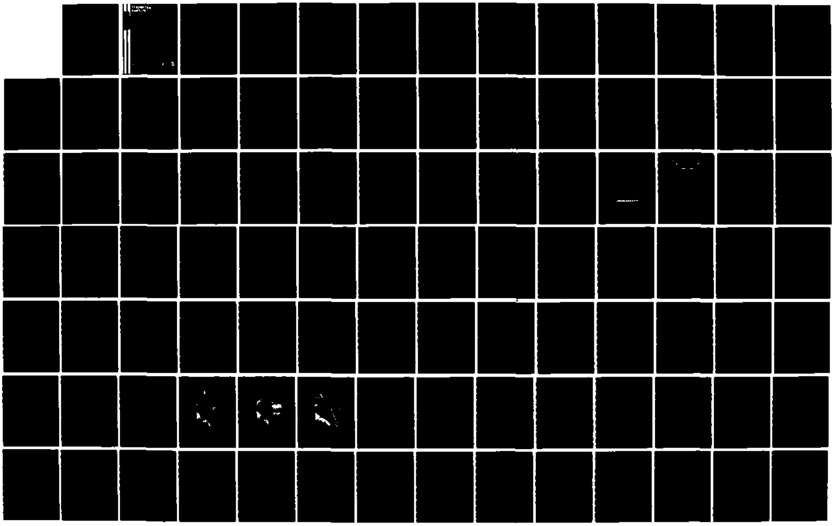
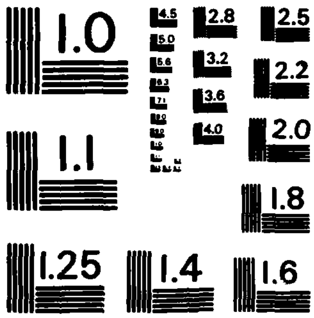
AD-A147 891 BASIC RESEARCH IN ELECTRONIC (JSEP): JOINT SERVICES 1/3
ELECTRONICS PROGRAM..(U) POLYTECHNIC INST OF NEW YORK
BROOKLYN MICROWAVE RESEARCH INST.. A A OLINER
UNCLASSIFIED 30 SEP 84 POLY-MRI-1432-83 AFOSR-TR-84-0978 F/G 9/5 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Polytechnic Institute of New York

AD-A147 891

AFOSR-TR- 84-0978

30 SEPTEMBER 1984

ANNUAL REPORT

on

BASIC RESEARCH IN ELECTRONICS (JSEP)

CONTRACT F49620-82-C-0084

APRIL 1, 1983 TO MARCH 31, 1984

DTIC FILE COPY

Approved for public release;
distribution unlimited.

DTIC
ELECTE
NOV 28 1984
S D
E

POLYTECHNIC INSTITUTE OF NEW YORK
MICROWAVE RESEARCH INSTITUTE
BROOKLYN, NEW YORK 11201

84 11 26 246

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 84 0978	2. GOVT ACCESSION NO. A147891	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Basic Research in Electronics (JSEP)		5. TYPE OF REPORT & PERIOD COVERED ANNUAL REPORT 1 April 1983-31 March 1984
		6. PERFORMING ORG. REPORT NUMBER POLY-MRI-1432-83
7. AUTHOR(s) Arthur A. Oliner		8. CONTRACT OR GRANT NUMBER(s) F49620-82-C-0084
9. PERFORMING ORGANIZATION NAME AND ADDRESS Microwave Research Institute Polytechnic Institute of New York Brooklyn, New York 11201		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F, 2305, A9
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NE Bolling Air Force Base Washington, D.C. 20332		12. REPORT DATE 30 September 1984
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Electromagnetics, microwaves, millimeter waves, waveguides and antennas, optics, x-rays, solid state interactions, solid state materials, surface acoustic waves, information electronics, systems, image restoration

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This Annual Report presents a summary of the scientific progress and accomplishments on research projects funded by the Joint Services Electronics Program (JSEP) for the contract period from 1 April 1983 through 31 March 1984. It does not contain information regarding accomplishments on research projects funded in other ways.

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.

The Joint Services Electronics Program at the Polytechnic is the core of interdisciplinary research in electronics encompassing programs in the Departments of Electrical Engineering and Physics under the aegis of the Microwave Research Institute. The research encompassed by this program is grouped under three broad categories: Electromagnetics, Solid State Electronics and Information Electronics. The detailed projects (work units) comprising the complete program are listed in the Table of Contents.

TABLE OF CONTENTS

Polytechnic Institute of New York
Microwave Research Institute

ABSTRACT

This Annual Report presents a summary of the scientific progress and accomplishments on research projects funded by the Joint Services Electronics Program (JSEP) for the contract period from 1 April 1983 through 31 March 1984. It does not contain information regarding accomplishments on research projects funded in other ways.

The Joint Services Electronics Program at the Polytechnic is the core of interdisciplinary research in electronics encompassing programs in the Departments of Electrical Engineering and Physics under the aegis of the Microwave Research Institute. The research encompassed by this program is grouped under three broad categories: Electromagnetics, Solid State Electronics and Information Electronics. The detailed projects (work units) comprising the complete program are listed in the Table of Contents.

SECTION I

ELECTROMAGNETICS

SECTION I: ELECTROMAGNETICS

A. NEW PHYSICAL EFFECTS INVOLVING OPEN DIELECTRIC STRUCTURES

Professors A.A. Oliner and S.T. Peng

Unit EM4-1

1. OBJECTIVE(S)

This study is concerned with guiding, radiating and scattering effects involving open dielectric structures. Almost all of the published literature on these topics involves a surface wave incident on a dielectric junction or discontinuity or grating, etc., at normal incidence, where the resulting boundary-value problem is two-dimensional, and TE and TM modes do not couple to each other. In connection with earlier investigations of open dielectric waveguides and antennas for millimeter-wave integrated circuits and for integrated optics, it was necessary for us to examine the behavior of dielectric structures for which the surface wave is obliquely incident. The boundary-value problems then change from the scalar two-dimensional ones to vector three-dimensional ones, and the TE and TM modes no longer remain independent but are coupled together. The vector nature of these problems introduces new mathematical challenges, but, more interestingly, we found that the mode coupling produced a rich variety of interesting and sometimes unexpected new physical effects.

In this contract, we continue to explore these physical effects on old and new types of dielectric structure. It is, of course, necessary to first obtain the appropriate mathematical solutions; then we examine the physical consequences of those solutions, and, where appropriate, assess their implications for device performance in millimeter-wave integrated circuits and integrated optics.

2. APPROACH

As mentioned in the section above, the analytical challenges involve vector three-dimensional problems which include TE-TM mode coupling. In addition, the presence of the TM modes produces very slow convergence, thus necessitating additional mathematical tricks to help accelerate the convergence. We have already produced a (novel) rigorous solution for a surface wave guided obliquely along a dielectric grating, by appropriately combining the solutions for two constituent scalar problems, and we have introduced a matrix partition procedure to speed up the mode convergence in the scattering by a dielectric step junction. All of the approaches we employ, both novel and standard, fall into the general category of guided wave techniques; we intend to continue to explore new techniques within this framework to obtain better solutions for existing problems and new solutions for new problems.

Since the basic goal in this study is the identification and understanding of new physical effects, it is best if we have available a rigorous solution, so that effects are not missed and spurious results do not arise due to approximations which are not fully understood. On the other hand, in most cases these rigorous results are quite complicated, requiring an involved computer program, particularly when complex roots are involved, as they often are. Therefore, after the physical effects are understood, and their implications for device performance assessed, we propose in some cases to derive simple approximate formulations, preferably in closed form, so that the magnitudes of these physical effects can be calculated quickly and easily. In addition, we expect to perform experiments at microwave or millimeter-wave frequencies to verify some of the physical effects.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contribution can be understood more clearly.

1) We were the first to show theoretically that certain dielectric strip waveguides for integrated optics or for millimeter wave integrated circuits can leak under appropriate circumstances, producing undesired cross talk between neighboring components. The only published experimental confirmation of that effect was indirect in nature. We have recently made measurements which constitute a direct experimental confirmation of these leakage effects, and we have presented a paper[1] at a symposium on this work. It is also interesting that some of the measurements were made in Japan by a collaboration we have established with a professor there; the paper was presented jointly with him.

2) Within the past few months, we have analyzed a new leaky wave structure in NRD (nonradiative dielectric) guide that may become a useful and novel leaky wave antenna for millimeter wavelengths. The behavior of the leakage constant as a function of guide width possesses features that strongly resemble those found for the various open dielectric waveguides we analyzed earlier. In particular, the sharp resonance dips for specific values of guide width represent an unusual performance feature that occurs here also. We have identified the reasons for these similarities, and we also note certain key differences. This study is continuing.

3) We had previously obtained careful experimental data from Professor R. Ulrich of the University of Hamburg-Harburg, West Germany, on dispersion values in the neighborhood of the Bragg reflection region for a periodic optical waveguide. That structure is composed of a periodic grating of dielectric grooves on a dielectric layer on a dielectric substrate, where the surface wave is incident at an angle on to the grooves, and the structure is of potential importance for multiplexers and other devices for integrated optics. When the surface wave is incident at an angle, as here, four stop bands are obtained in the dispersion plot instead of the usual two. We earlier developed a rigorous theory for such situations, and we previously compared our theory with Ulrich's measurements and found excellent agreement in general shape and even reasonable agreement in numbers. Curves showing both the measurements and theory were presented previously, but we believed that the discrepancy arose because our theory applied to rectangular grooves, whereas their measurements were made on grooves which were more sinusoidal than rectangular. Somewhat over a year ago, we found out that the dimension given to us earlier by letter for the layer thickness was the value before the grooves were etched, and that an SEM

photo of the structure showed that the grooves were primarily rectangular, not sinusoidal. When we recalculated our theoretical values using the new value for layer thickness, remarkable agreement with the earlier measurements was obtained.

4) One of the important new physical effects introduced by oblique, rather than normal, surface wave incidence on a periodically-grooved dielectric layer is that of cross polarization, due to the TE-TM wave interactions occurring at every groove step when the incidence angle is oblique. Studies of cross polarization were begun earlier, but during this past year we have completed a study of such cross polarization when the periodic structure is designed to radiate. We expected a significant amount of cross polarization to be present, but we were actually surprised by the strength of it. For normal incidence on the grooves there is no cross polarization, of course, but the amount becomes quite strong even for small deviations from normal. For an incident TE surface wave, for example, the TM content of the radiated power actually exceeds the TE content for angles from the normal greater than 15° or so. For an incident TM surface wave, even stronger cross polarization effects are found; for example, in one case, for an angle somewhat less than 15°, a maximum occurs at which about 90% of the radiated power consists of the "wrong" polarization.

5) Within this past year, we have derived a simple new method for the analysis of periodically grooved dielectric image guide antennas of narrow width. The method yields the values of the phase and leakage constants for antennas of arbitrary width in terms of (simpler) known expressions for "equivalent" antennas of infinite width. An effective dielectric constant approach is employed to establish the equivalence. Very good numerical agreement was found with measurements on an antenna only $0.11\lambda_0$ wide; these measurements were supplied by the U.S. Army at Ft. Monmouth, NJ.

6) When a number of these new physical effects are placed together, they form a cohesive and very interesting package. With this viewpoint in mind, we prepared an overview paper which we presented at a symposium[2] in August 1983. The presentation was very well received, with many compliments. It was later accepted for publication.[3] Another such package, containing only material relating to optical frequencies, has also been accepted for publication.[4]

7) The investigations so far have concentrated on effects involving dielectric strip waveguides and periodically grooved dielectric layers with obliquely incident waves. Although such investigations are not complete, we have also begun an examination of a somewhat more complicated structure: a dielectric layer grooved periodically in two orthogonal directions, resembling a waffle iron. The reason for the new investigation is that our earlier brief look into the behavior of such a structure indicated that an enormous variety of weird-looking wavenumber diagrams may be possible. The implication is that additional new physical effects may emerge from such a study. That look was performed "kinematically," i.e., by simply examining various possibilities graphically, employing the various unperturbed interacting wavenumber curves. Recently, we have begun an analytical investigation, taking a model in which the variations in one direction are independent of those in the other orthogonal direction. During this past year, we have examined

analytically the behavior of an intermediate stage, that for a wave propagating at an arbitrary angle in the periodic medium itself. Even that calculation is very involved and is not available in the literature. We have obtained dispersion curves for several cases, and we found an unexpected and interesting result. It turns out that selection rules exist which govern when crossings in the dispersion plot couple together or simply remain as crossings. This study has complicated facets and is just beginning.

4. STATE OF THE ART AND PROGRESS DETAILS

As the earlier studies progressed, it became clear that the methods of analysis and, in many cases, the basic physical behavior of dielectric waveguides for integrated optics and for millimeter-wave integrated circuits are quite similar and have much in common, although important minor differences persist. In addition, it became evident that before the operation of these waveguides and antennas could be understood completely it was necessary to solve certain basic canonical guiding and scattering problems involving open dielectric structures. The solutions in the literature for these canonical problems were either nonexistent or only approximate. Our attempts to solve these problems revealed that the problems were vector in nature, whereas almost everything previously available was only scalar, and that the TE and TM modes necessarily became coupled together. In the course of the analyses, we were fortunate enough to recognize, where others did not, that the TE-TM coupling gives rise to a variety of interesting physical effects, some of them unexpected, and that these effects have implications for device performance.

A. Guidance and Leakage Properties of Dielectric Strip Waveguides

Examples of dielectric strip waveguides for integrated optics and for millimeter-wave integrated circuits (almost all of which we have analyzed) are shown in Fig. 1. Dielectric strip waveguides may be viewed as composed of two dielectric step junctions with a length of planar dielectric waveguide between them. Propagation along such waveguides may then be described in terms of a surface wave which bounces back and forth between the sides, or step junctions, undergoing "total reflection" at each bounce.

Let us first recognize that the approximate theories, such as the Marcatili procedure[5] and the effective dielectric constant method,[6,7] assume that only one surface wave mode is present and that the geometrical discontinuities at the step junctions can be neglected. These approximate methods, which have been widely applied[6,8-11] to the structures in Fig. 1, thus neglect the TE-TM mode coupling occurring at the sides of the waveguides, and they thus miss the interesting physical effects. Such coupling produces a leaky mode instead of the purely bound modes predicted by the approximate theories.

The situation that gives rise to leaky modes is summarized in Fig. 2. Without the TE-TM mode conversion at the sides of the waveguide, only the TE mode is considered to be present by the approximate theories. That assumption requires that the incident and reflected TE "rays" in Fig. 2 are above cut-off (propagating), whereas the transmitted "ray," in the outside region, is below cut-off (evanescent), and that no TM

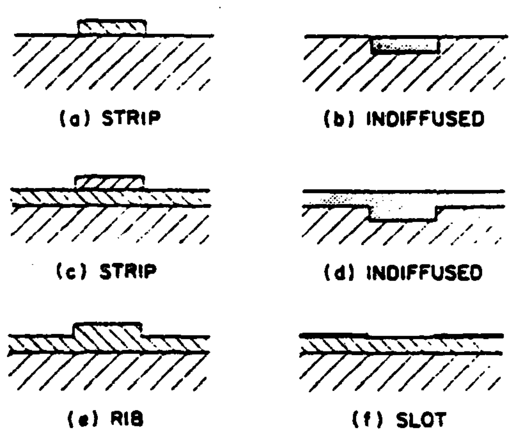


Fig. 1(a) Examples of open dielectric waveguides for integrated optics.

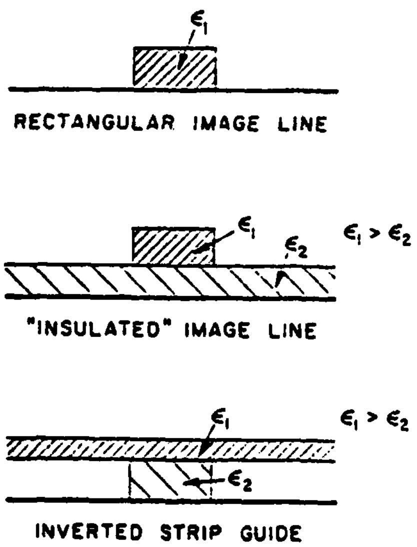


Fig. 1(b) Examples of open dielectric waveguides for millimeter waves.

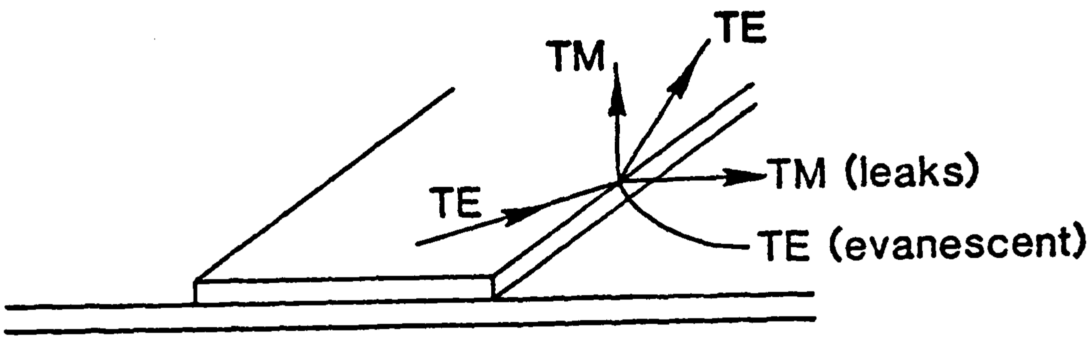


Fig. 2 Pictorial representation of mode-conversion effects at the side of a dielectric waveguide that give rise to leaky modes.

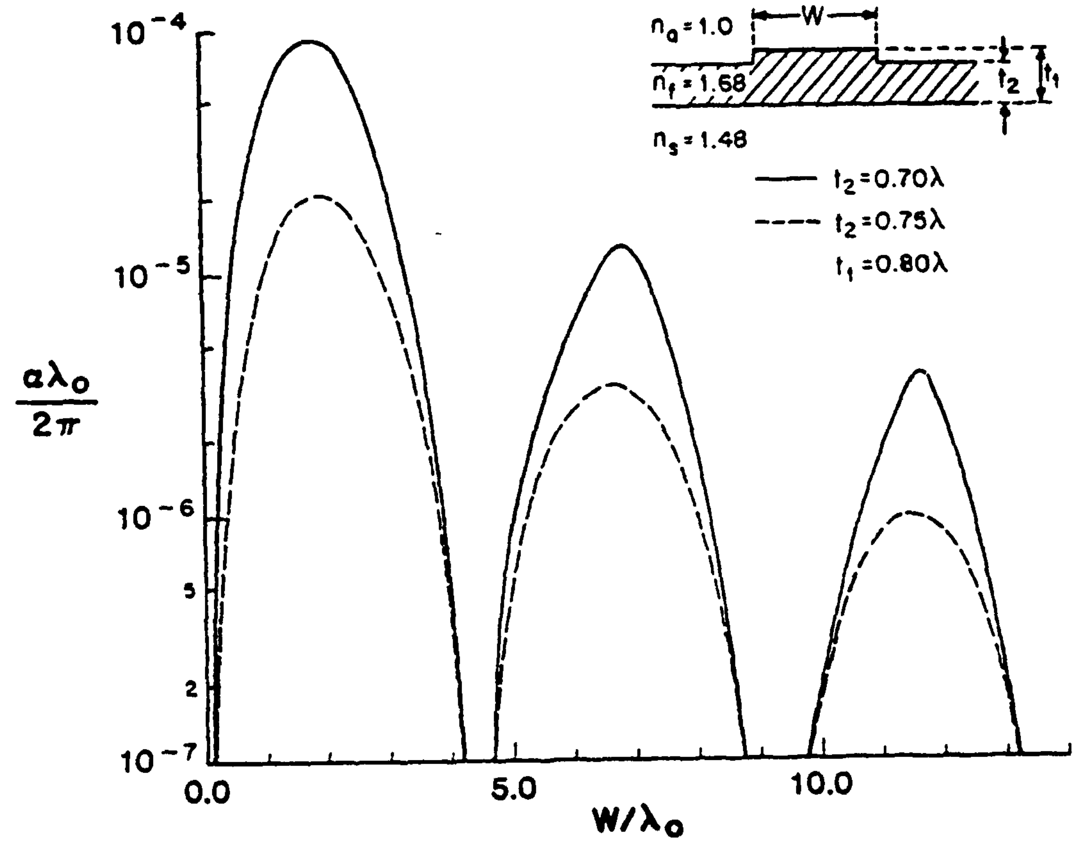


Fig. 3 Plot of attenuation constant α as a function of guide width W for a rib waveguide for integrated optics. Note the periodic sharp resonance dips.

mode contributions are present. In the more accurate situation, where mode coupling is included, TM mode "rays" must be added, as seen in Fig. 2. Since we are depicting the situation for which the TM mode is dominant, the reflected TM wave inside is certainly propagating but the transmitted TM wave outside may be propagating or evanescent, depending on the geometrical parameters. In most cases, that wave will be propagating; when it is, as shown in Fig. 2, it indicates that energy in the TM polarization will be leaking per unit length along the waveguide, with the result that the propagation constant is complex, with a non-zero attenuation constant α.

It is important to note that the dominant portion of the guided wave energy possesses TE polarization but that the leakage energy has TM polarization, in the form of a TM surface wave propagating away from the waveguide at some angle. In all other known cases of leaky modes, the leakage energy possesses the same polarization as the exciting energy. The class of leaky modes described here thus constitutes a new class of such modes.

A typical plot of the attenuation constant α of the leaky mode as a function of the guide width W is shown in Fig. 3. For a purely bound mode, the value of α would be identically zero (since we are neglecting material losses here). The non-zero value of α is thus due directly to the TM wave outside in Fig. 2. We observe that the curve in Fig. 3 also shows periodic large dips, which are "resonances" or cancellation effects. These strong dips are due to the mode-converted TM wave in the inside region of Fig. 2, which also bounces back and forth between the sides of the waveguide. Thus, the mode-converted transmitted wave produces the leakage, and the mode-converted reflected wave produces the resonance dips in Figure 3. These are severe physical effects, not minor ones, and they are completely un-expected on the basis of the earlier published theories, which neglected the mode-conversion effects.

We have made calculations on most of the waveguides shown in Fig. 1. We have shown on fundamental grounds[12,13] why the dielectric image guide for millimeter waves, and guides (a) and (b) for integrated optics, never leak, and why most modes on the remaining waveguides will ordinarily leak. We have also found that the magnitude of the leakage may be large or small, depending on the type of waveguide, and we have shown why. We have also presented[12,13] specific detailed examples for several of the waveguides shown in Fig. 1. For example, the leakage is generally small for waveguides for integrated optics, and for the inverted strip waveguide for millimeter waves, whereas it can become very large under appropriate circumstances for the insular guide. For a special case of the insular guide, the attenuation was as large as 4 dB/λ!

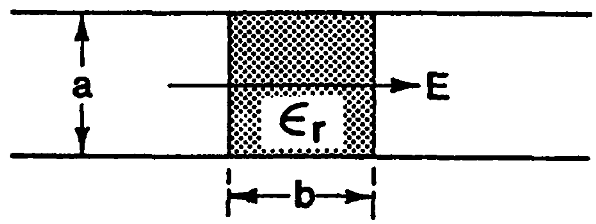
We were gratified to note a publication[14] by Japanese authors that describes an experiment they performed to detect leakage from a rib-type dielectric waveguiding structure. They make strong reference to our theoretical work and our physical explanation for the effect, and they conclude that their measurements verify the theoretical predictions.

A basic difficulty, or limitation, associated with that experiment is that it is indirect. The leakage corresponds only to one constituent of the total insertion loss, and the measurement simply shows that the total insertion loss becomes modified in a manner consistent with our theoretical predictions when certain parameters are varied. We therefore set up a measurement program of our own in which the leaky wave and its properties would be measured directly by careful probing procedures.

In connection with measurements of this type, we established a collaboration this past year with Professor H. Shigesawa and his colleagues at Doshisha University, Kyoto, Japan. Our measurements were taken at a wavelength of about 3 cm, whereas those in Japan were taken at a wavelength of 6 mm (f = 50 GHz), so that a greater waveguide length could be probed well. In addition, those measurements in Japan demonstrated clearly that no radiation was measureable in the resonance regions where sharp dips in α were obtained theoretically. A joint paper that combined our measured results with theirs was presented at a symposium.[1]

Within the past few months, we have analyzed a new leaky wave structure in NRD (nonradiative dielectric) waveguide which may lend itself as a useful and novel antenna for millimeter wavelengths. Ordinary NRD guide and the new leaky wave structure based on it are shown in Figs. 4(a) and (b), respectively.

(a)



(b)

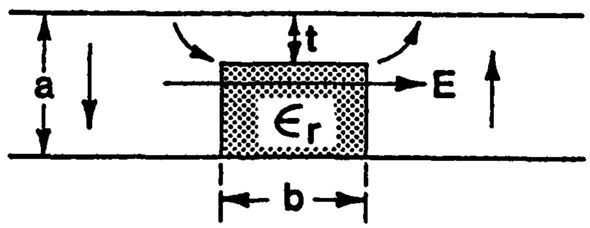


Fig. 4 (a) NRD (nonradiative dielectric) guide, which is like H guide (rotated through 90°), except that spacing a is made < $\lambda_0/2$ to insure that discontinuities are reactive.
(b) New leaky wave structure based on NRD guide.

NRD guide is a variant of the well-known H guide, which was proposed about 20 years ago as a low-loss waveguide for millimeter waves. In order to reduce the loss even further, the spacing between the plates was made greater than a half wavelength; although the loss was lowered, radiation was produced at every bend or discontinuity along the guide, so that components were hard to build. The H guide was therefore impractical. Recently, Yoneyama and Nishida,[15] from Japan, proposed that the spacing be reduced to less than a half wavelength, with the result that all bends and discontinuities became purely reactive. These authors also showed how components could be built easily in that waveguide, and they proceeded to design and build many of them.[16,17]

In connection with a contract[18] with the Air Force at Hanscom Field, we analyzed and measured a novel leaky wave antenna based on NRD guide. That antenna is very different from the structure in Fig. 4(b), however, being based on the full height of the dielectric strip; leakage was produced there by foreshortening one side of the guide. During the course of the measurements, we found that an air gap sometimes developed between the dielectric strip and the upper metal plate. We overcame the difficulty by employing a very thin double-backed adhesive strip, but the leaky wave structure of Fig. 4(b) resulted from speculations about what effect such an air gap would have.

In the NRD guide of Fig. 4(a), the electric field is parallel to the plates, and the fields are bound to the region of the dielectric strip, being evanescent away from the strip in the air regions outside. Any bends or discontinuities would excite the TE_1 mode in parallel plate guide in the air region, but that mode is below cutoff; as a result, those bends or discontinuities would be purely reactive. When an air gap is present, however, as in Fig. 4(b), the asymmetry also excites the TEM mode in the air region; since that mode is always above cutoff, we could reasonably expect leakage to occur in the form of that TEM mode.

After a bit of thought about how to analyze the structure of Fig. 4(b), we recognized that this structure is a member of the family of dielectric strip waveguides that we have been discussing above, and that the general computer program that we have developed for this class of guides could be adapted for this special case with only minor modifications. Accordingly, we made calculations of the leakage constant α as a function of the dielectric strip width b, with the air gap thickness t as a parameter.

Some typical numerical results for the leakage constant α are shown in Fig. 5. The dielectric strip width b in Fig. 5 is the same as W in Fig. 3; we use b here because that symbol is customarily employed for that dimension in NRD guide. The first thing to note, if we compare Figs. 5 and 3, is the remarkable similarity in shape, especially the presence of the sharp dips. This strong similarity exists despite certain important differences between the dielectric strip guides and the structure in Fig. 4(b).

In the dielectric strip guides, the TE (or TM) constituent surface wave excites the mode-converted TM (or TE) surface wave as it strikes the side of the dielectric strip. The mode-converted TM (or TE) surface wave in the outside region produces the leakage, whereas the component of that wave bouncing back and forth inside the dielectric strip region causes the sharp dips in the leakage constant α for certain specific guide

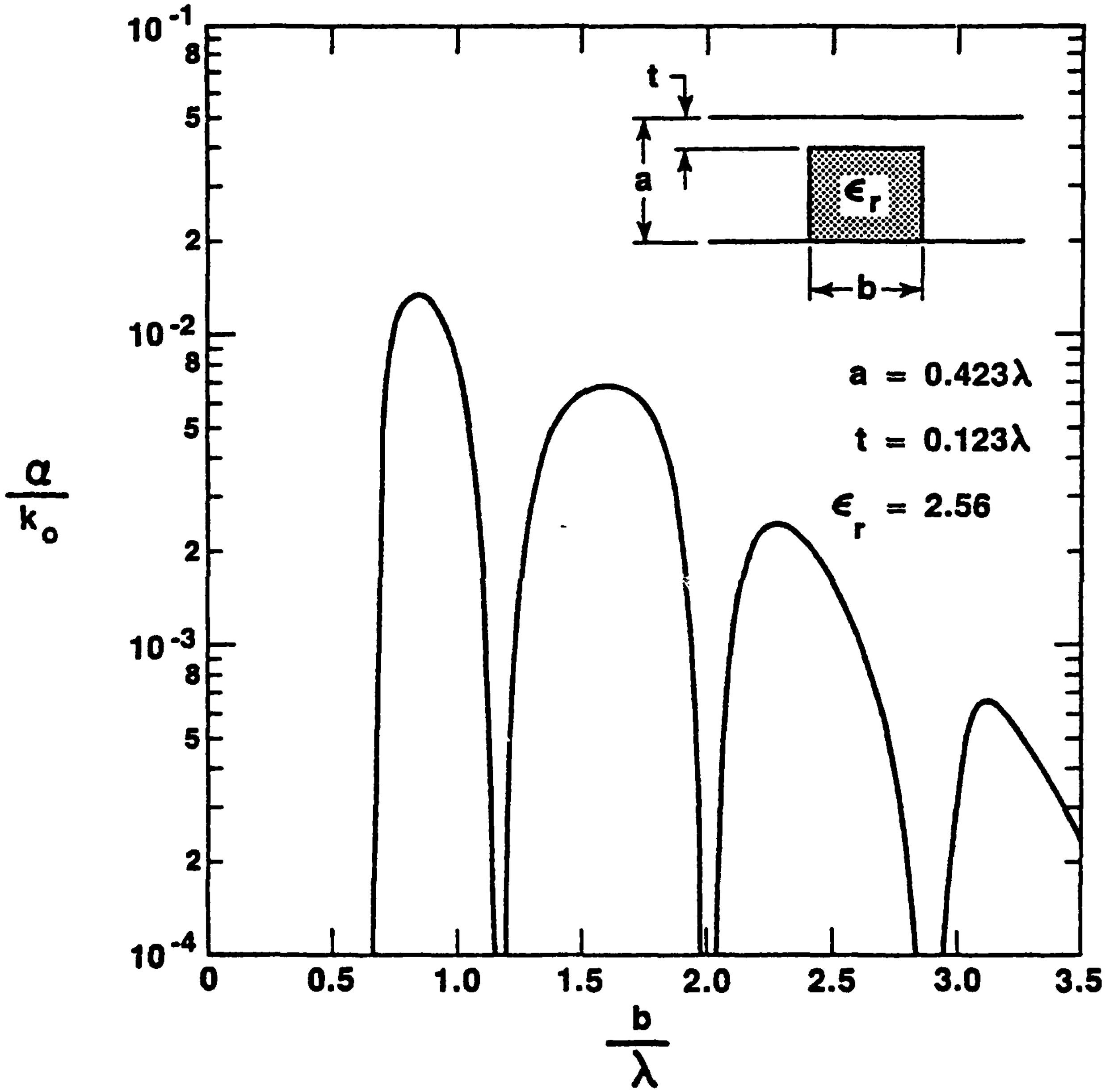


Fig. 5 Attenuation constant α as a function of dielectric strip width b for the new leaky wave structure in NRD guide shown in Fig. 4(b). Note the similarity to Fig. 3.

widths. In the structure of Fig. 4(b), the TE constituent surface wave, when it strikes the side of the dielectric strip, excites the mode-converted TM surface wave within the dielectric strip as usual, but also the TEM mode in the air region outside of it. In this case, the mode-converted TM surface wave does not exist in the air region, and it is the TEM mode that carries the power that leaks away. Thus, the α is caused by the TEM mode outside, but the resonant dips in the α values relate to the TM surface wave inside.

Another important difference is present between the two structures. In the dielectric strip guides, the outside region can always support a surface wave in some direction, so that its effective dielectric constant for the basic mode is real and positive. In contrast, the outside region in Fig. 4(b), corresponding simply to the parallel plates, is below cutoff for the basic mode (the TE_1 mode), so that its effective dielectric constant in that region is negative real. As a result, the lowest longitudinal guided mode for the overall structure is above cutoff for <u>any</u> value of the guide width W in the case of the dielectric strip guides, whereas, for the structure in Fig. 4(b), the longitudinal guided mode will go <u>below</u> cutoff if the width W (or b) becomes sufficiently small.

This feature explains why the curve for α in Fig. 3 exists down to W = 0, whereas the curve in Fig. 5 stops at a non-zero value of width b. There are some other interesting differences in the behaviors of the two structures which are still being examined.

B. <u>Oblique Guidance by a Planar Periodically-Grooved Dielectric Layer</u>

This study began as a first step in the rigorous analysis of radiation from periodically-grooved grating antennas on millimeter-wave dielectric antennas, but we soon realized that the solution is useful and applicable in its own right in integrated optics. The study would then be that of optical surface waves propagating at an angle on a periodically-corrugated dielectric layer on a dielectric substrate. Such structures, with obliquely-incident surface waves, have been finding increasing application lately in such components as multiplexers, filters and mode deflectors.

The guidance of optical surface waves propagating in a direction <u>normal</u> to the grooves on a periodically-corrugated planar dielectric waveguide is well known, and various theoretical treatments are available, some of the most accurate having been developed at our Institute. Such structures have been used in grating couplers and in distributed-feedback lasers, for example. The electromagnetic boundary-value problem for the case of normal incidence is scalar, however, where the TE and TM surface waves remain independent of each other. When the surface propagates at an <u>oblique</u> angle with respect to the grooves, the TE and TM surface waves are now coupled together, and the boundary-value problem becomes a three-dimensional vector one. The problem is one for which an exact solution has not been available previously.

About four years ago, we derived a <u>rigorous</u> analytical solution to the vector, oblique-incidence problem which is valid for all ranges of parameter values; we believe that this solution represents the first exact analysis of a three-dimensional vector guided-wave problem involving a periodic structure.

SECTION I: ELECTROMAGNETICS

A rich variety of physical effects arise as a result of oblique guidance, which are not present at all for normal guidance. One immediate consequence, due to the TE-TM coupling which occurs at every groove step, is that of polarization conversion to the other mode type; that is, if a TE mode is incident, some energy would be mode-converted into a TM mode. Another effect, explained below, is that at stop bands, corresponding to Bragg reflection, additional stop bands appear. Although many other interesting effects can arise, these two effects were recognized previously in two papers[19,20] which also presented approximate theoretical treatments valid only for small periodic perturbations, and only when the waves are purely bound. Our rigorous analysis is of course valid for rectangular grooves of arbitrary depth, aspect ratio or period.

Suppose we consider a grooved planar dielectric waveguide with the grooves parallel to the y direction, and the x direction perpendicular to them. The propagation behavior of waves guided at an oblique angle to these grooves is represented by a wavenumber plot, β_y vs. β_x. Such a plot indicates the behavior as a function of angle with respect to the grooves, but at a fixed frequency. Alternatively, a k vs. β plot presents the behavior as a function of frequency for a given angle. In both of these plots, the periodicity in the x direction, which causes space harmonics to be excited, produces additional curves on these plots repetitive in x. Where certain of these curves cross, stop bands are present, the first of these stop bands being known also as Bragg reflection.

Separate curves are present for each possible mode, and then for each space harmonic of each mode. Usually, the waveguide thickness is such that only one TE surface wave mode and one TM surface wave mode are present. If we then consider the Bragg reflection condition, the stop band corresponds to the crossing of the n=0 (basic) wave and its n=-1 space harmonic. When no coupling exists between the TE and TM surface waves, we obtain one TE and one TM Bragg stop band. However, when oblique guidance occurs, the TE and TM surface waves do couple, and four stop bands are possible instead of only two. The two additional stop bands correspond respectively to the n=0 TE and n=-1 TM curves, and to the n=0 TM and n=-1 TE curves, couplings which could not occur for normal guidance.

We were fortunate in learning about the availability of accurate experimental data in the above-mentioned stop-band regions. We first learned of these measurements at a Workshop,[21] and then further at an Optical Society Topical Meeting.[22] The speaker, Dr. Reinhard Ulrich of Germany, indicated that no accurate theoretical data were available with which he could compare his measurements, and he was delighted to learn of our theory. Dr. Ulrich provided us with detailed data, and we developed a computer program for the required numerical values.

A comparison between our theory and his measurements showed excellent qualitative agreement, with stop bands wide where they should be, etc. We had understood that the grooves on which the measurements were made were more sinusoidal than rectangular; since our theory applies to rectangular grooves, we did not expect that the comparison would be numerically exact. More recently, however, we learned that the dimension given to us earlier by letter for the dielectric layer thickness was the value before the grooves were etched. The etching process also reduced the layer thickness somewhat; although the final thickness was not known, a good estimate was available for the probable range of the reduction in thickness. In addition, we learned from the Ph.D. student, R. Zengerle, who was familiar with the intimate details, that an SEM photo of the structure had shown that the grooves were primarily rectangular, not sinusoidal as we were informed initially.

We therefore recalculated our theoretical results by employing a more correct value for the layer thickness. We took the middle of the probable range for the reduction, which then reduced the layer thickness from 0.165 μm to 0.159 μm. This small change produced a dramatic result. The theoretical values now agree quantitatively with the measurements to a remarkable degree. This new comparison is shown in Fig. 6, where it should be noted that the region shown is an enlargement of the area of the wavenumber plot just in the neighborhood of the stop bands.

Various physical implications follow from these wavenumber curves. First, we recognize that in this type of plot the phase-velocity direction is given by the line from the origin, whereas the group-velocity or energy-flow direction is given by the perpendicular to the actual curve. Thus, in the vicinity of the stop bands there is substantial anisotropy; that is, the energy-flow direction can be quite different from the direction of phase progression. This effect is called beam steering.

Another set of interesting physical effects occurs when the frequency is raised sufficiently for one or more of the space harmonics to become radiating. The first space harmonic to radiate is the $n = -1$ space harmonic, and it will radiate first into the substrate. For a given incidence angle into the periodic grating, one can determine how many beams will radiate and where. If one examines carefully the exit angle of the $n = -1$ space harmonic, one finds very peculiar results. In general, the beam radiates at some skew angle with respect to the plane of incidence. We have not as yet examined this skew effect carefully, but we can see that the potential exists for unusual opportunities.

An additional new physical effect that arises because of the TE-TM mode coupling produced by oblique guidance is cross polarization. Again, this effect is absent entirely when the guidance is perpendicular to the grooves. The amount of cross polarization cannot be assessed without employing the complete dynamic theory, which we have done recently.

We therefore extended the basic theory to encompass this new direction. Our previous results permitted us to compute the propagation characteristics of these surface waves guided obliquely on a planar grooved dielectric waveguide for integrated optics. Such computations led to dispersion curves for the behavior of these waves. Mathematically phrased, the solutions obtained yielded the eigenvalues for these guided surface waves. The theory was now extended to include the eigenvectors; that is, the solutions now also include the relative amplitudes of all the field components. The theory now permits us to determine the polarization content of the complicated fields that result when the TE and the TM constitutent surface waves become coupled at each step of the grating.

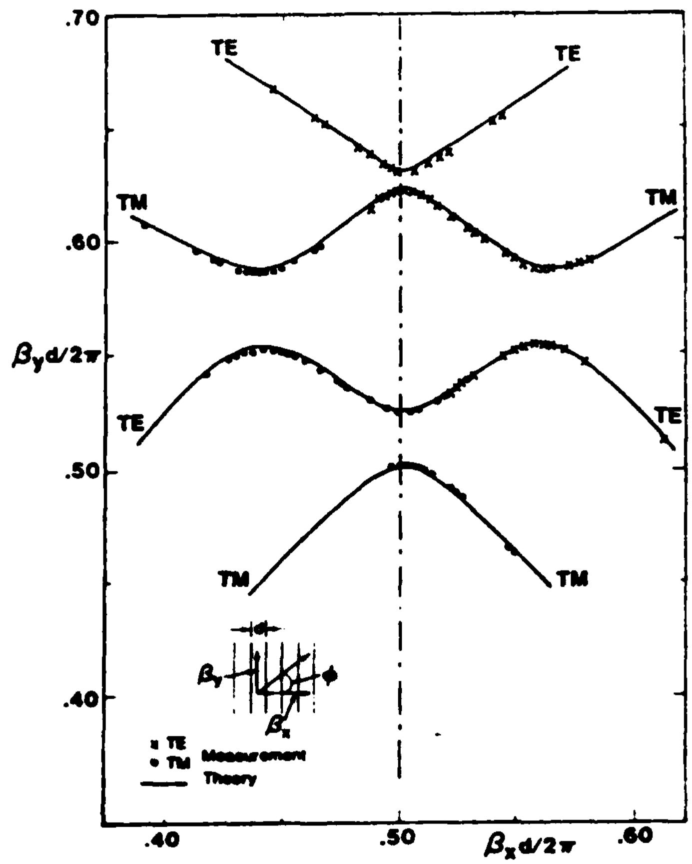


Fig. 6 Bragg interaction region for periodic optical waveguide. Comparison between rigorous theoretical calculations (solid line) and accurate measurements (points) tak[illegible]y Ulrich and Zengerle[22] for the stop band region of the wavenumber diagram for surface waves propagating at an oblique angle ϕ to a periodically grooved dielectric layer on a dielectric substrate. Note that the ordinate and abscissa scales are greatly expanded, corresponding to enlargement of the wavenumber diagram just in the neighborhood of the stop bands. The top and bottom stop bands, centered at $\beta_x d/2\pi = 0.50$, are due to TE-TE and TM-TM interactions; the TE-TM stop bands are shifted from the customary 0.50 value. The various parameters are: groove thickness $t_g = 0.048$ µm, film plus groove (total layer) thickness $t_f + t_g = 0.159$ µm, period $d = 0.282$ µm, free space wavelength $\lambda_o = 0.6117$ µm, refractive indices of film n_f and substrate n_s are, respectively, 2.10 and 1.47.

We have applied this extension of the theory to the case of groove spacing wide enough to permit the n=-1 space harmonic to become radiating; all the other space harmonics remain bound, however. When the surface wave is guided in the direction normal to the grooves, the polarization of the radiating space harmonic remains simple, the same as that of the incident surface wave. As that surface wave is rotated with respect to the grooves, corresponding to oblique incidence, the polarization content becomes mixed.

We expected a significant amount of cross polarization to be present, but we were actually surprised by the strength of it. For normal incidence on the grooves there is no cross polarization, of course, but the amount becomes quite strong even for small deviations from normal. For an incident TE surface wave, for example, the TM content of the radiated power actually exceeds the TE content for angles from the normal greater than 15° or so. For an incident TM surface wave, even stronger cross polarization effects are found; for example, in one case, for an angle somewhat less than 15°, a maximum occurs at which about 90% of the radiated power consists of the "wrong" polarization. Some early results were presented at a symposium[23] in May 1983, and these studies of polarization content also form part of a recent Ph.D. thesis.[24]

Some numerical results are summarized in Figs. 7 and 8, which apply respectively to TE and to TM incident surface waves. For the TE case, for example, we see in Fig. 7 that there is no vertical (z) component of electric field at normal incidence ($\phi = 0$); only a vertical magnetic field is present (normalized to permit comparison in amplitude with the electric field component). The curves show that even for small deviations from the normal the cross-polarization content builds up quickly, and the surprising effects mentioned above occur in both Figs. 7 and 8.

C. Dielectric Gratings of Finite Width, and Implications for Device Performance

Dielectric waveguides for millimeter-wave integrated circuits possess narrow widths so as to permit single-mode operation. Many people have proposed that gratings be placed on such waveguides to achieve two types of device: resonators and filters, and leaky-wave antennas. For the resonators and filters, the gratings are operated in the Bragg reflection region so as to achieve strong reflections combined with negligible radiation. For the antenna application, the gratings are operated in the leaky-wave region so as to achieve controlled radiation with negligible reflection.

Since wave guidance by these open dielectric waveguides may be viewed in terms of a pair of surface waves propagating at an angle to the waveguide axial direction, the new physical effects discussed under B above apply here as well when a grating is placed on the waveguide. This situation is illustrated in Fig. 9 when a grating is placed on a dielectric image guide. Some differences are present, of course, between this situation and that of oblique guidance on an infinitely-wide grating, in that here a pair of waves, rather than a single wave, must be considered, and the influence of end effects must be included.

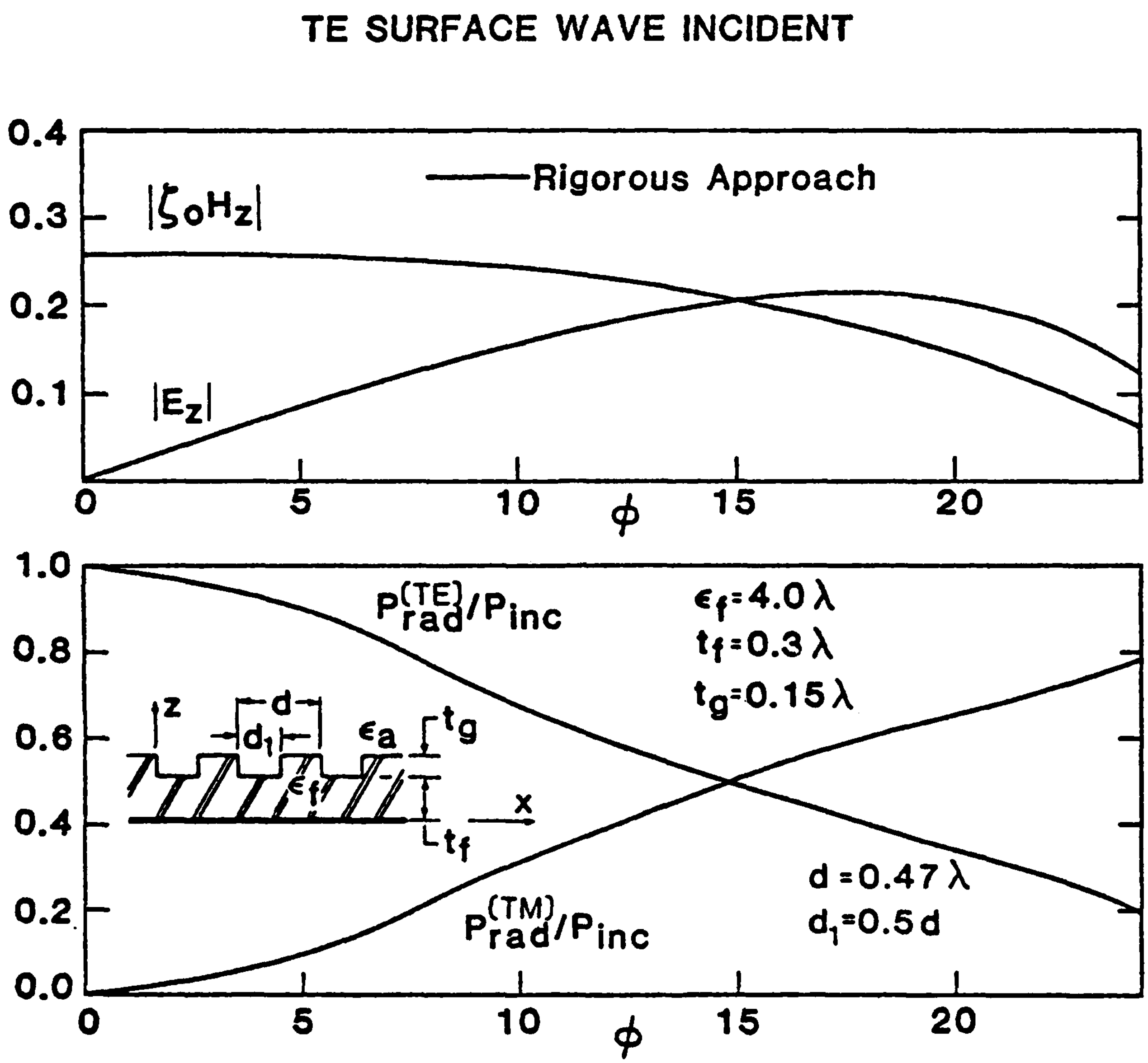


Fig. 7 The direct and the cross-polarized radiation field components due to a surface wave obliquely incident on a periodically grooved dielectric layer, showing the surprisingly large amount of cross polarization. The upper curves present the field components perpendicular to the air-dielectric interface; the lower curves indicate the radiated fields. For TE surface wave incident.

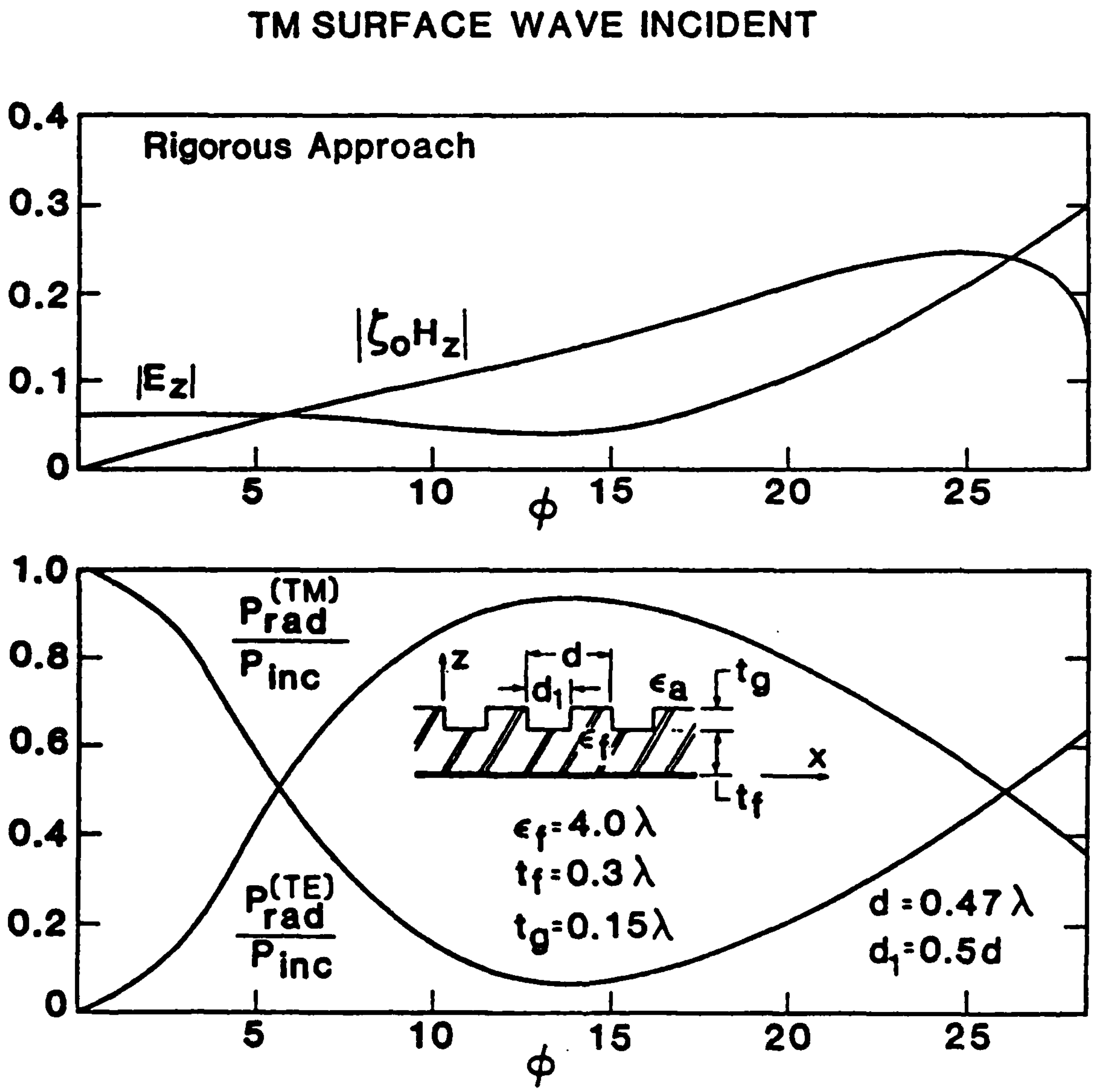


Fig. 8 Same as for Fig. 7, except for TM surface wave incident.

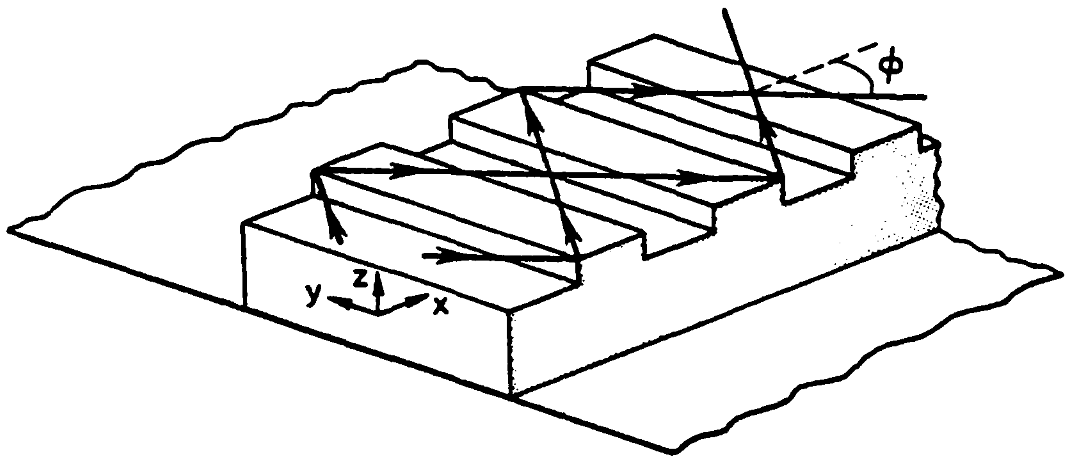


Fig. 9 The guidance of surface waves along a periodically grooved dielectric image guide, showing constituent waves propagating at an angle.

Our investigation of the effect of finite width on resonator or filter performance has shown that when the grating of finite width is operated in the Bragg regime an extra (unwanted) transmission dip or reflection peak may be obtained, which may seriously impair the device performance. The extra reflection peak near to the desired peak arises from an extra stop band due to interaction between TE and TM space harmonics, and in fact has a polarization opposite to that in the desired peak. This effect thus corresponds to the presence of the extra stop bands shown in Fig. 6, which were due to such TE-TM coupling. No such problem was encountered with SAW resonators, where the grooves are very wide, and it was assumed that narrowing the width would not introduce any new problems. That supposition can be quite wrong because a new physical effect is introduced as a result of the finite width, as we pointed out.[25]

It is worth noting that Japanese authors[20] have devised a multiplexer for integrated optics that utilizes two periodic gratings at an angle to the incident surface wave. They found experimentally that an extra reflection peak, of opposite polarization, arises in this process due to the TE-TM coupling.

Many measurements have been taken on leaky-wave antennas comprised of gratings on dielectric image guide, of the type shown in Fig. 9. No theory for the leaky-wave behavior had appeared, however, taking finite width into account, until our recent presentations.[26,27] Our theory adapts the analytical result mentioned earlier for oblique guidance by a grating and presents a straightforward procedure for determining the α and β of the leaky wave, thus permitting one to design such an antenna in accordance with desired radiation characteristics. Ours is the only theory available to date that yields such information.

That design suffers in only one way, however. It is based on the rigorous theory for oblique guidance by a grooved dielectric surface, which is rather involved mathematically and therefore has an associated complex computer program. A practical design procedure should require a less involved computer program. Toward that end, we derived during the past year an alternative and simpler procedure.

This new solution makes use of an effective dielectric constant to transform the antenna of narrow width into an equivalent one of infinite width, so that the previously obtained formula[28] for the phase and leakage constants for antennas of infinite width can be readily utilized. In employing this effective dielectric constant concept, the width and the dielectric constant are changed to compensate each other, but all other physical and structural parameters are maintained constant.

This study was conducted in conjunction with the U.S. Army at Ft. Monmouth, New Jersey, and they supplied the measurements against which the new theoretical procedure was checked. The structure on which the measurements were made was only $0.11\lambda_0$ wide, truly a narrow antenna. We made theoretical calculations for a variety of cases, including the measured antenna structure, by employing this simple approach, namely, by obtaining the equivalent leaky wave structure of infinite width and then utilizing the already available expressions for α.

A comparison is presented in Fig. 10 between such theoretical values and the measurements referred to above, where the leakage constant is plotted as a function of frequency in the millimeter wave range. It is evident that the agreement is rather good, except in the vicinity of the sharp dip and rise in the theoretical curve. That region corresponds to broadside radiation, at which a leaky wave stop band occurs. The agreement is not expected to be good there since the sharp behavior predicted by the theoretical structure of infinite length would be greatly softened by the finite length of the practical structure on which the measurements were made.

This procedure also has its limitations, which should be understood. It is valid for certain integrated properties, such as the phase and the leakage constants, but it cannot provide information on some field properties that vary as a function of position, such as cross-polarization features away from the principal plane. For the radiation properties in the principal plane, however, this method is not only simple but highly useful.

This new simple procedure, together with the comparison with experiment, is being presented in June 1984 at a symposium,[29] with the U.S. Army collaborators as coauthors.

When a number of these new physical effects are placed together, they form a cohesive and very interesting package. The first of such packages corresponded to a keynote invited paper;[30] leakage from dielectric waveguides was also included in that one. A second such package, involving only topics from subsections B and C, was presented last fall,[2] and was accepted for publication.[3] A third such package, containing only material relating to optical frequencies, has also been accepted for publication.[4]

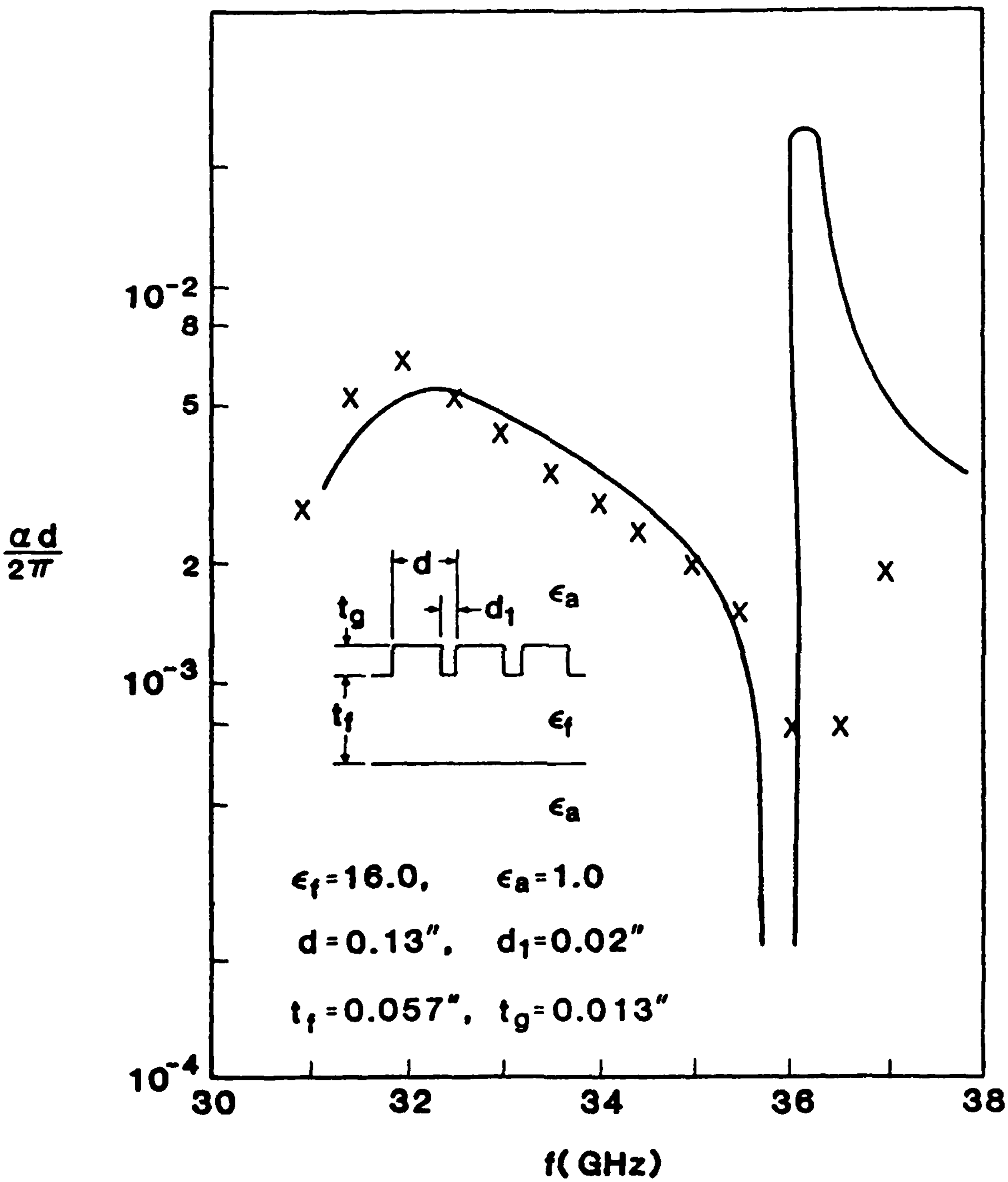


Fig. 10 Comparison between theoretical values (solid line) and experimental results (crosses) on a millimeter wave antenna of width $0.11\lambda_0$.

D. Two-Dimensionally Periodic Goooved Dielectric Structures

In subsection B, we described the very interesting variety of new physical effects that one finds when a surface wave is incident obliquely on a dielectric waveguide that is grooved periodically in one dimension. Those physical effects, which are absent when the incidence is normal instead of oblique, include extra stop bands due to TE-TM mode coupling, anisotropy effects such as beam steering, radiation at skew angles, and strong cross-polarization effects. Our calculations were also shown to agree extremely well with careful measurements made by Ulrich and Zengerle on an optical guiding structure.

Ulrich and Zengerle, in this paper[22] and especially in Zengerle's Ph.D. thesis,[31] of which we have a copy, also took careful measurements on optical waveguides which were grooved periodically in two orthogonal directions. They found additional interesting physical effects, such as weird-looking dispersion diagrams, and the focusing of a divergent light beam. These were determined experimentally; they did not have any theory available, but they attempted to explain the results qualitatively by using kinematic (as opposed to dynamic) considerations.

The possibility of finding additional physical effects, and the availability of the Ulrich-Zengerle measurements, were the motivating factors in this new study of a two-dimensionally periodic grooved dielectric waveguide. The development of the two-dimensionally periodic structure from the original smooth planar waveguide, applied to integrated optics, is shown in Fig. 11 and is taken from Fig. I.1.1 on page 19 of Zengerle's thesis.[31] As seen, the final structure resembles a waffle iron. If the one-dimensionally periodic structure of Fig. 11(b) is now known to permit added flexibilities due to the new physical effects that result, it is reasonable to speculate that the two-dimensionally periodic one of Fig. 11(c) may offer even more.

The investigation of the behavior of the two-dimensionally periodic waveguide was begun recently. The dispersion relation for the guidance properties will be obtained by taking the transverse resonance in the vertical (z) direction (see Fig. 11), as was done for the one-dimensionally periodic case. As an intermediate first step, we need the propagation behavior for a wave propagating at an arbitrary angle in the periodic medium itself, that is, when height w in Fig. 11 is made infinite. Even that calculation is very involved, and is not available in the literature. In addition, it is necessary to select a mathematical model for the variation of the dielectric constant in the x and y directions.

Our procedure is therefore the following. First, we select the above-mentioned mathematical model for the dielectric constant variation of the structure with no variation in the z direction. Then, we solve for the propagation behavior when the wave propagates at an arbitrary angle in the xy plane but without variation with z. Finally, to achieve the full intermediate solution, we include the variation with z, corresponding to a wave in the infinite medium with arbitrary propagation angle. Then, we mode match at the various interface boundaries to obtain the result that applies to Fig. 11(c).

So far, we have performed only the first stage, but we have already found some interesting results. We chose a dielectric constant

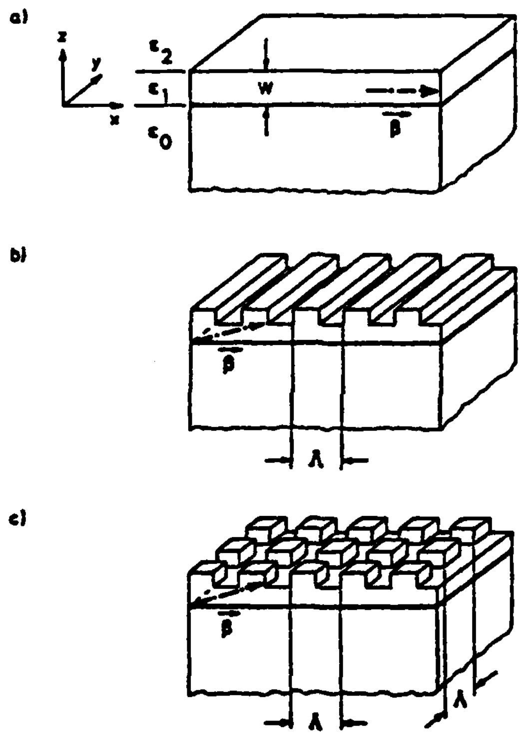


Fig. 11 The development of a two-dimensionally grooved periodic dielectric waveguide from the original smooth waveguide. (a) Original smooth waveguide. (b) One-dimensionally grooved structure. (c) Final two-dimensionally grooved structure. (Taken from Zengerle's Dr.-Ing. thesis.[31])

with independent x and y variations, but in product form, which seems reasonable:

$$\varepsilon(x,y) = \varepsilon_{av.} \left[1 + 4\delta \cos \frac{2\pi}{a} x \cos \frac{2\pi}{b} y \right] \tag{1}$$

where δ is the modulation index and a and b are the grating periods in the x and y directions, respectively. If the fields are assumed to follow the product form for no variation with z, and a is set equal to b, the wave equation separates into two, one for the x+y coordinate and one for the x-y coordinate. These two separate wave equations, which are related by a mutual unknown constant, are each satisfied by Mathieu functions in view of the cosine dielectric constant variation. For this special case, therefore, we have obtained separability in a neat fashion, and we can obtain numerical results in terms of canonical functions (the Mathieu functions) which are well understood.

From the two Mathieu differential equations, we obtain two dispersion relations for x+y and x-y separately, or for x and y in a coupled sense. Each dispersion relation can be expressed in terms of rapidly convergent continued fractions;[32] after eliminating the mutual constant, a dispersion relation for the wavenumbers k_x and k_y is obtained.

A new, unexpected, and interesting result has been obtained from this study so far. It turns out that <u>selection rules</u> exist which control when crossings in the dispersion plot couple together or simply remain as crossings. The rule is the following: Those crossings that correspond to couplings require that the space harmonic numbers of one curve must differ from <u>both</u> of those of the other curve. Numerical calculations for the wavenumber plots have been obtained for several parameter values; two are shown in Figs. 12 and 13. In both of these examples, a square lattice is taken (a = b), with modulation index $\delta = 0.10$. In the former case, for $ka/2\pi = a/\lambda = 0.6$, only simple crossings are obtained. The space harmonic values for the separate curves that cross are seen to differ from each other in either x or y, but not both. In the second case, shown in Fig. 13, additional crossings are possible and the wavenumber plot becomes much more complicated. It is seen that now crossings occur between curves whose space harmonic values differ in <u>both</u> the x and y directions, such as the (0,0) and the (-1,-1) curves, and that those crossings <u>couple</u>, producing stop bands.

These interesting results associated with this early stage of the study of two-dimensionally periodic dielectric structures are being presented at a symposium[33] in June 1984.

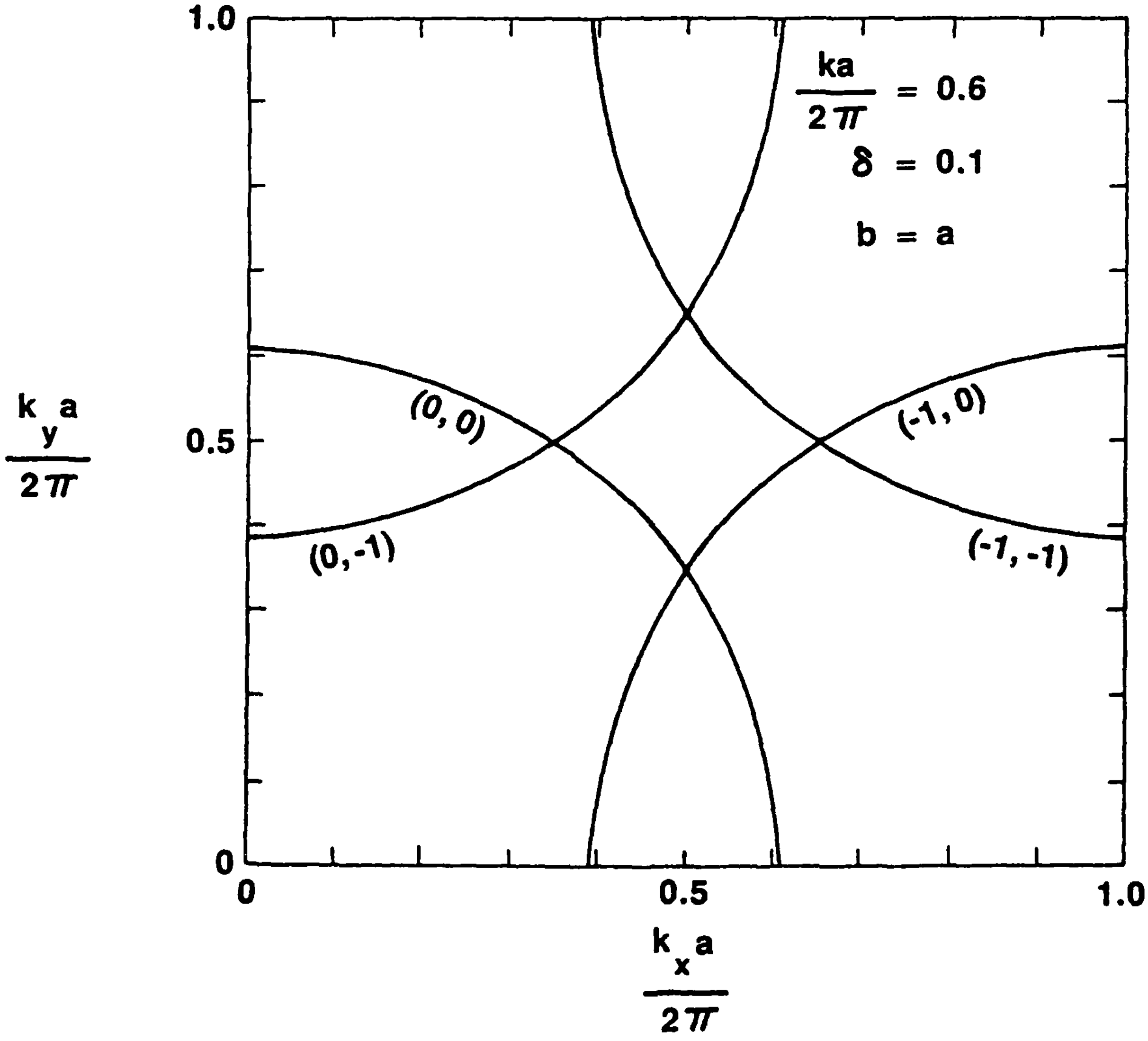


Fig. 12 Wavenumber plot for infinite medium with dielectric constant modulated sinusoidally in two orthogonal dimensions. Case for equal periods and for $ka/2\pi$=0.6, showing only crossings but no stop bands.

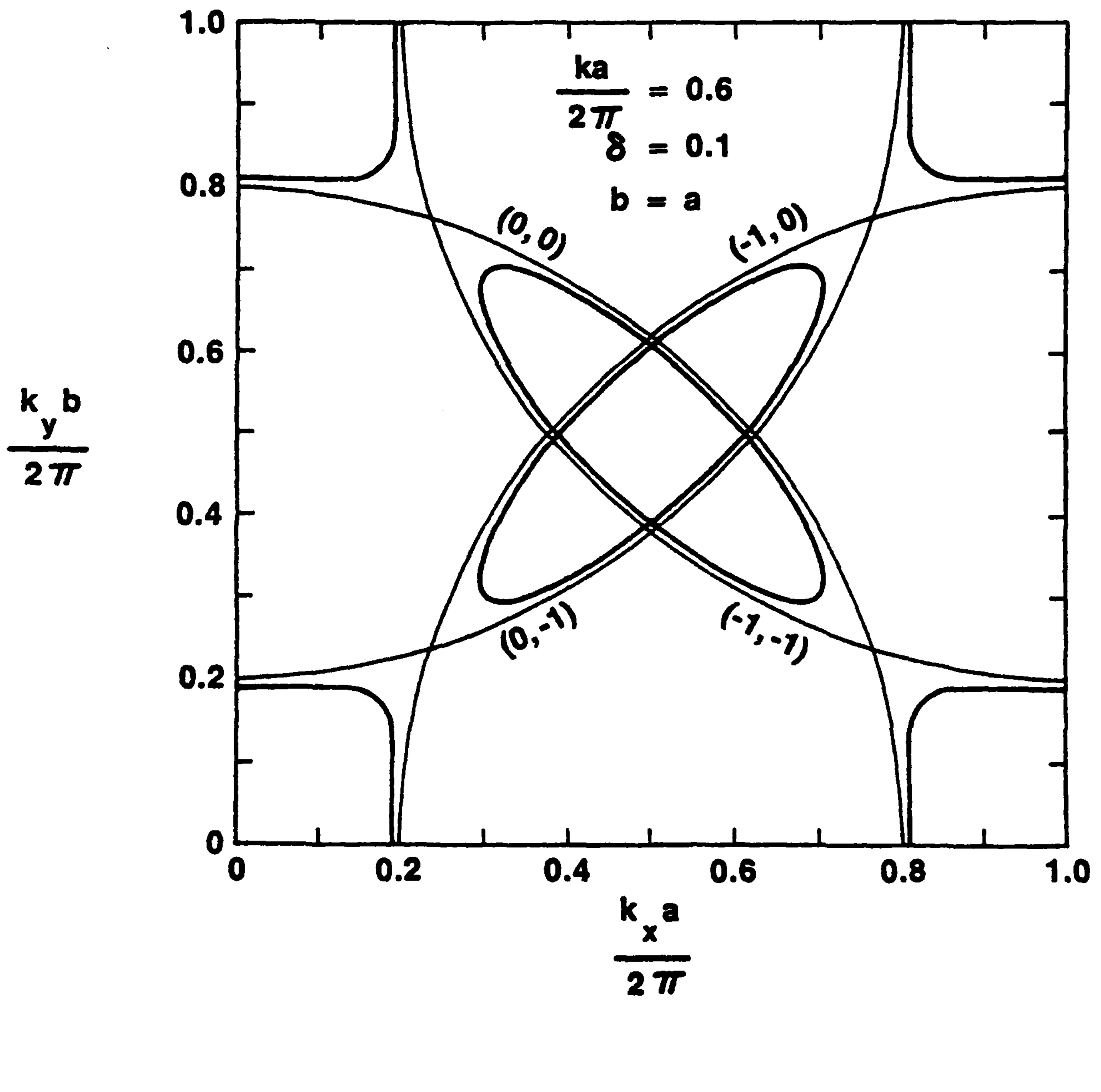


Fig. 13 Same as Fig. 12 except for $ka/2\pi=0.8$, showing that some stop bands appear when the frequency or the period is increased.

5. REFERENCES

1. H. Shigesawa, M. Tsuji, J.S. Myung, S.T. Peng and A.A. Oliner, "Direct Experimental Confirmation of New Leakage Effects on Open Dielectric Strip Waveguides," Digest of IEEE International Microwave Symposium, pp. 293-295, Boston, Mass. (May 31 - June 3, 1983).

2. A.A. Oliner and S.T. Peng, "New Physical Effects on Periodically-Grooved Open Dielectric Waveguides," Proc. URSI International Symposium on Electromagnetic Theory, pp. 515-518, Santiago de Compostela, Spain (August 23-26, 1983).

3. A.A. Oliner and S.T. Peng, "New Physical Effects on Periodically Grooved Open Dielectric Waveguides," Radio Science, accepted for publication.

4. A.A. Oliner and S.T. Peng, "New Physical Effects on Periodically Grooved Optical Planar Waveguides," Applied Scientific Research, accepted for publication.

5. E.A.J. Marcatili, "Dielectric Rectangular Waveguide and Directional Coupler for Integrated Optics," Bell System Tech. J., Vol. 48, pp. 2071-2102 (September 1969).

6. W.V. McLevige, I. Itoh and R. Mittra, "New Waveguide Structures for Millimeter-Wave and Optical Integrated Circuits," IEEE Trans. Microwave Theory Tech., Vol. MTT-23, pp. 788-794 (October 1975).

7. R.M. Knox and P.P. Toulios, "Integrated Circuits for the Millimeter Wave Through Optical Frequency Range," Proc. Sympos. on Submillimeter Waves, Polytechnic Press of Polytechnic Institute of Brooklyn, pp. 497-516 (April 1970).

8. H. Furuta, H. Noda and A. Ihaya, "Novel Optical Waveguide for Integrated Optics," Appl. Opt., Vol. 13, pp. 322-326 (February 1974).

9. V. Ramaswamy, "Strip-Loaded Film Waveguides," Bell System Tech. J., Vol. 53, pp. 697-704 (April 1974).

10. N. Uchida, "Optical Waveguide Loaded with High Refractive-Index Strip Film," Appl. Opt., Vol. 15, pp. 179-182 (January 1976).

11. T. Itoh, "Inverted Strip Dielectric Waveguide for Millimeter-Wave Integrated Circuits," IEEE Trans. Microwave Theory Tech., Vol. MTT-24, pp. 821-827 (November 1976).

12. S.T. Peng and A.A. Oliner, "Guidance and Leakage Properties of a Class of Open Dielectric Waveguides, Part I: Mathematical Formulations," IEEE Trans. Microwave Theory Tech. Vol. MTT-29 (Special Issue on Open Guided Wave Structures), pp. 843-855, September 1981. Invited Paper.

13. A.A. Oliner, S.T. Peng, T.I. Hsu and A. Sanchez, "Guidance and Leakage Properties of a Class of Open Dielectric Waveguides, Part II: New Physical Effects," same as reference 12. Invited Paper.

14. K. Ogusu and I. Tanaka, "Optical Strip Waveguide: An Experiment," Appl. Opt., vol. 19, pp. 3322-3325 (1 October 1980).

15. T. Yoneyama and S. Nishida, "Nonradiative Dielectric Waveguide for Millimeter-Wave Integrated Circuits," IEEE Trans. on Microwave Theory Tech., Vol. MTT-29, No. 11, pp. 1188-1192 (November 1981).

16. T. Yoneyama and S. Nishida, "Nonradiative Dielectric Waveguide Circuit Components," International Conference on Infrared and Millimeter Waves, Miami, Florida (December 1981).

17. T. Yoneyama, F. Kuroki and S. Nishida, "Design of Nonradiative Dielectric Waveguide Filter," Digest IEEE International Microwave Symposium, pp. 243-244, San Francisco, CA (May 30 - June 1, 1984).

18. Contract No. F19628-81-K-0044, with Rome Air Development Center, Hanscom Field, MA.

19. A. Gudzenko, "Bragg Reflection in Planar Dielectric Waveguides with Periodic Thickness Modulation," Radio Engineering and Electron Physics (USSR), Vol. 22, pp. 19-25 (1976).

20. K. Wagatsuma, H. Sakaki and S. Saito, "Mode Conversion and Optical Filtering of Obliquely Incident Waves in Corrugated Waveguide Filters," IEEE J. Quantum Electronics, Vol. QE-15, pp. 632-637 (July 1979).

21. Fourth Workshop on Optical Waveguide Theory, Noordwijkerhout, The Netherlands (September 1979).

22. R. Ulrich and R. Zengerle, "Optical Bloch Waves in Periodic Planar Waveguides," Topical Meeting on Integrated and Guided Wave Optics, Incline Village, Nevada (January 28-30, 1980).

23. M.J. Shiau, S.T. Peng and A.A. Oliner, "Strong Polarization Conversion in Radiation from Surface Waves Incident Obliquely on a Grooved Dielectric Layer," Digest of National Radio Science Meeting, p. 103, Houston, Texas (May 23-26, 1983).

24. M.J. Shiau, "Scattering and Guidance of Electromagnetic Waves by Periodic Dielectric Structures," Ph.D. Thesis, Polytechnic Institute of New York (June 1983).

25. M.J. Shiau, H. Shigesawa, S.T. Peng and A.A. Oliner, "Mode Conversion Effects in Bragg Reflection from Periodic Grooves in Rectangular Dielectric Image Guide," Digest International Microwave Sympos., pp. 14-16, Los Angeles, California (June 15-17, 1981).

26. S.T. Peng, A.A. Oliner and F. Schwering, "Theory of Dielectric Grating Antennas of Finite Width," Proc. of IEEE AP-S Internat. Sympos., pp. 529-532, Los Angeles, California (June 16-19, 1981).

27. S.T. Peng and A.A. Oliner, "Radiation from Grating Antennas on Dielectric Waveguides of Finite Width," Proc. 11th European Microwave Conference, Paper No. B8.5, Amsterdam, The Netherlands (September 7-11, 1981).

28. F. Schwering and S.T. Peng, "Design of Dielectric Grating Antennas for Millimeter-Wave Applications," IEEE Trans. Microwave Theory Tech., Vol. MTT-31, pp. 199-209 (April 1983).

29. S.T. Peng, M.J. Shiau, A.A. Oliner, J. Borowick, W. Bayha and F. Schwering, "A Simple Analysis Procedure for Dielectric Grating Antennas of Finite Width," Proc. IEEE International Symposium on Antennas and Propagation, Boston, MA (June 25-29, 1984).

30. A.A. Oliner, "New Physical Effects on Open Dielectric Waveguides Due to TE-TM Mode Coupling," Proc. Seventh Colloquium on Microwave Communication, pp. 390-397, Budapest, Hungary (September 6-10, 1982). Invited paper.

31. R. Zengerle, "Lichtausbreitung in ebenen periodischen Wellenleitern," Dr.-Ing. Thesis, Stuttgart University, West Germany (1979).

32. T. Tamir, H.C. Wang and A.A. Oliner, "Wave Propagation in Sinusoidally Stratified Dielectric Media," IEEE Trans. Microwave Theory Tech., Vol. MTT-12, No. 3, pp. 323-325 (May 1964).

33. S.T. Peng and T.L. Dong, "Wave Propagation in Two-Dimensionally Periodic Medium," Proc. IEEE International Symposium on Antennas and Propagation, Boston, MA (June 25-29, 1984).

SECTION I: ELECTROMAGNETICS

B. WAVE INTERACTIONS AND ABSORPTION RESONANCES ON OPEN LOSSY STRUCTURES

Professor T. Tamir

Unit EM4-2

1. OBJECTIVE(S)

To explore basic aspects of the scattering, guiding, and absorption of electromagnetic waves by stratified and/or periodic media having intrinsic losses. Unlike previously studied structures that were assumed to be ideally lossless, the present situations are characterized by physical parameters that include realistic absorption, conversion, or radiation losses. In particular, the investigation will focus on the recently observed phenomenon of anomalously high energy absorption that occurs under circumstances involving only slight losses.[1-9] For this purpose, the projected study will use simplified canonic configurations that characterize a wide variety of situations and thus facilitate the clarification of fundamental concepts. Specifically, the results obtained would be relevant to integrated-optical applications, to selective spectral filters for electromagnetic or optical purposes, to photo-detectors, to the guiding by and leakage from geological structures, as well as to analogous acoustic-wave situations involving propagation along planar interfaces.

2. APPROACH

The wave interactions and the resonant-absorption behavior of canonic configurations with realistic losses are studied by utilizing rigorous techniques[16,17] that have already been developed to a high degree of sophistication and success in the area of microwave engineering and in integrated optics.[15] In this context, extensive use is made of results and procedures already available from some of our earlier studies.[18-23]

In addition to well-established techniques, we use a novel approach that applies complex pole-zero analysis to the parameters describing scattering and guidance. As an illustration, the reflectance of stratified media is a well-defined known analytic function which fully describes the reflection, transmission, and guidance properties in terms of the poles and zeros located in the wavenumber plane. The tracking of the pertinent pole-zero loci in that plane can therefore serve both for the analysis and for the synthesis of structures having prescribed behavior. All of these methods are employed analytically, with subsequent verification by numerical computer methods[16] and possibly also by experimental observations.[9]

An important aspect of the program is that the pertinent wave interactions are studied by considering first the scattering of incident plane waves, which yield relatively simple but idealized descriptions of the pertinent process. However, the plane-wave results are then extended to the scattering of incident realistic (e.g., Gaussian) beams, which provide an accurate physical description of actual situations. The extension to the beam-wave case is carried out by using integrations in the complex wavenumber plane, which are evaluated by suitable asymptotic techniques and other analytical or numerical approaches.[17-23]

3. STATE OF THE ART AND PROGRESS DETAILS

The propagation, scattering, and guiding of waves by planar configurations have been studied extensively in multilayered[10] and periodic[11] structures. However, most of these studies have examined ideally lossless configurations. The presence of losses has usually been treated as a perturbation on the ideal models, so that its effect was pronounced only if propagation occurred over very large distances. Thus, for small intrinsic losses, the scattered field is usually very similar to that of the lossless case, and the absorbed energy is then only a small percentage of the incident energy. As an example, for the simple case of a layered structure placed on a perfect conductor, as shown in Fig. 1(a), the reflectance ρ has a magnitude of unity. If small intrinsic losses occur in one of the layers, $|\rho|$ is generally close to unity for layers having a thickness t of the order of the wavelength λ, because the power flux then traverses a relatively small path inside the lossy medium, as suggested by the first-order refracted and reflected ray in Figure 1(a). A similar situation holds for periodic structures of the type shown in Figure 1(b). However, under critical combinations of the incidence angle and the physical parameters, the fields are strikingly different in that <u>all</u> of the incident energy is absorbed and the scattered waves are then suppressed.

This anomalous-absorption behavior was first identified in periodic structures[1,2,4] by the French group led by Professor Petit, but we have subsequently shown[3,7-9] that such a phenomenon occurs in a more general class of planar configurations, which may involve only a few layered media. A similar absorption behavior has been studied[12-14] in optics in the context of the so-called "induced transmission" through thin films. However, that transmission effect was restricted to normal incidence on stratified media, whereas the anomalous absorption phenomena to be explored occur at generally oblique incidence angles and they appear also in periodic structures.

Our earlier studies have revealed[3,5,9] that, by taking a leaky-wave approach to the scattering phenomenon, it is possible to describe the anomalous effects in terms of a strong interactive coupling of energy between the incident wave and a leaky-wave field supportable by the structure. This interaction produces an electromagnetic-field regime whereby the energy flux travels over a lengthy longitudinal path D along the structure, as suggested in Figure 2. Thus, even for a layer having a thickness t of the order of the wavelength λ and very small losses per λ, the overall attenuation of the energy flowing along $D \gg \lambda$ accounts for the anomalously high absorption effect. In this context, our previous studies have also demonstrated the important result that this effect can occur for both types of polarization,[3,7] and that it takes place in a wide class of multilayered structures.[9] In addition, we have experimentally verified[8] that the anomalous effect, which had previously been observed only in metallic gratings,[1,2] can nevertheless occur also in simple multilayered configurations.

The above discussion has focused on the anomalous-absorption effect because of its peculiar and interesting aspects. However, it must be emphasized that this effect is only one facet of the wave-interaction processes that occur in lossy configurations. As discussed further below, a larger class of wave-interaction phenomena (mostly of the leaky-wave type) represents the broader scope of the present investigation. Because the various wave interactions are well covered by describing the progress made during the past period of the present contract, we list below details of our studies and their more important results.

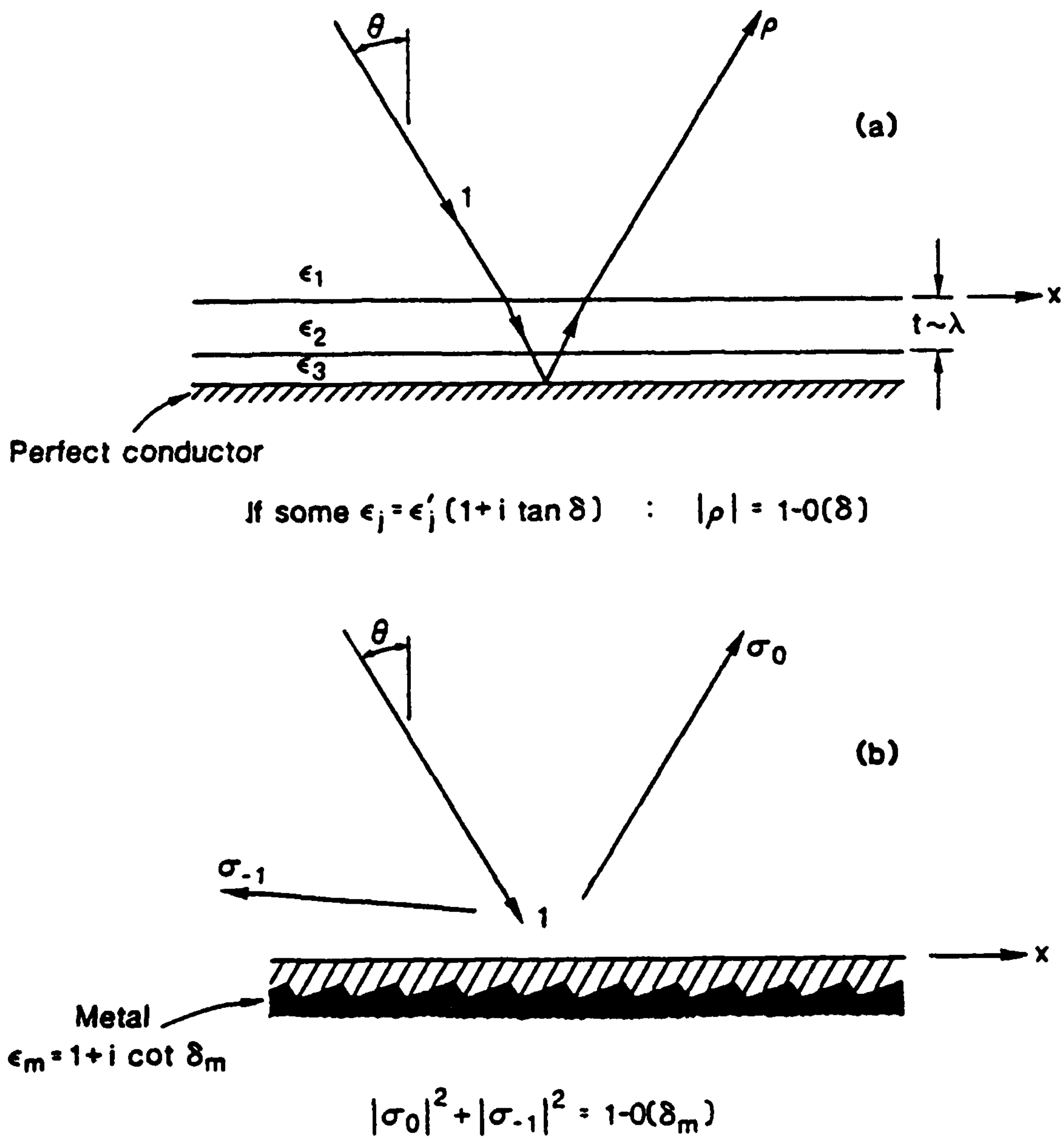
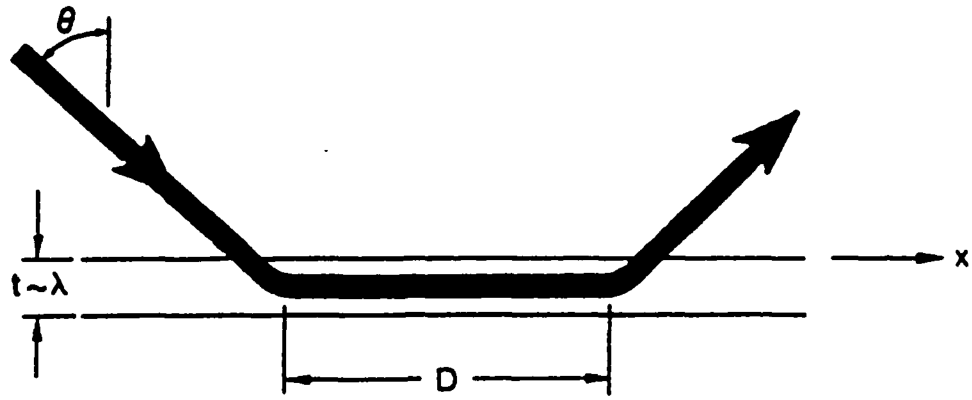


Fig. 1 Scattering by thin planar structures. (a) Incidence and reflection at a layered configuration, (b) diffraction by a grating with only two propagating orders.



Leaky-wave field: $E \sim e^{ik_{xp}x} = e^{i(\beta_p + i\alpha_p)x}$

If $k \sin\theta = \beta_p$ and $\alpha_p \ll \beta_p$, then

$$D \simeq \frac{2}{\alpha_p} \gg \lambda$$

Fig. 2 Leaky-wave interpretation of resonant absorption, showing how the incident energy is trapped into the structure in the form of a longitudinal power flux that flows along a distance D after which the energy is re-radiated back into the upper medium.

(a) Most of the work on the anomalous absorption effect and related phenomena has been carried out by examining leaky waves of the TE type.[24-32] In this context, it was known for a long time that, for a layer placed between two semi-infinite media, those waves leak into both exterior media and they are labeled as air leaky waves. However, recent work[24,27-31] has revealed that a novel type of leaky wave can occur which radiates into only the denser of the two exterior media; these waves are labeled as substrate leaky waves. We have therefore carried out an in-depth study of the novel (substrate) waves and have explored their relation to the previously known (air) variety of leaky waves, as well as to surface waves. In particular, we have determined how all of these waves evolve as frequency varies. The results of this detailed analytic study are being published in a forthcoming paper.[33] The relevance of the novel type of leaky waves is mentioned in item (c) below.

(b) Because the work mentioned under (a) above was restricted to TE modes, we have initiated a study of the analogous waves for TM modes. Preliminary results show that, as frequency varies, the evolution of the TM modes is quite different from that for TE modes. While the behavior of the air leaky waves is only quantitatively different for TE and TM leaky modes, the substrate leaky waves show also qualitative differences. In particular, TM substrate leaky modes fall into three different categories that are differentiated from each other by the presence of Brewster conditions, which do not exist for TE modes. As work on this topic is still continuing, we expect to describe further results in our subsequent reports.

(c) Most of the past work on the scattering of Gaussian beams by planar structures[7,18,19] has dealt with basically reflecting configurations, i.e., structures in which the transmitted beam is absent or negligible. We have therefore explored the scattering of Gaussian beams by basically transmitting configurations, for which the transmitted and reflected beams have comparable intensities. We have found that, in these configurations, both the air and substrate types of leaky waves described in item (a) above may affect the profile of the reflected and transmitted beams. In particular, we have shown that these beams can be significantly displaced from the position predicted by geometrical-optics considerations and their profiles may be significantly distorted from the Gaussian shape. Furthermore, the reflected beam can undergo an unexpected backward shift under certain easily obtainable conditions. These results have been reported at meetings,[29 31] a brief article has recently been published,[33] and a comprehensive paper is being submitted for publication.[34]

(d) To gain further insight into the physical processes involved in the effects mentioned above, i.e., anomalous absorption, beam displacements, leaky-wave phenomena, etc., we have initiated a new study dealing with the power flux of beams rather than their field quantities. In this approach, we track the local Poynting vector and thus attempt to understand the actual trajectory of the beam energy as it is reflected from, or transmitted through, layered media. Because this necessitates the use of complex and highly accurate computer programs, we have acquired a Hewlett-Packard 300 computer system which is capable of providing the required data in a reasonably short time. An example of preliminary results that we have recently obtained is shown in Fig. 3 for a beam incident on a slab of thickness $h=5\lambda$ and dielectric constant $\varepsilon_1=3.0$, placed between an upper (air) region and a lower (substrate) region, with dielectric constants $\varepsilon_0=1.0$ and $\varepsilon_2=1.5$, respectively. It is interesting to note that the reflected beam is derived from the right-hand portion of the incident beam flux, while the transmitted beam is produced by the left-hand portion of that flux. This description helps to explain the general distortion of the outgoing beams as well as their lateral displacements. These results are scheduled to be presented at the next annual meeting of the Optical Society of America.[36] A systematic study of these power flux aspects is scheduled for the next period.

4. REFERENCES

1. M.C. Hutley and D. Maystre, "The Total Absorption of Light by a Diffraction Grating," Optics Commun., Vol. 19, pp. 431-436 (December 1976).

2. E.G. Loewen and M. Nevière, "Dielectric Coated Gratings: A Curious Property," Appl. Optics, Vol. 16, pp. 3009-3011 (November 1977).

3. V. Shah and T. Tamir, "Brewster Phenomena in Lossy Structures," Optics Commun., Vol. 23, pp. 113-117 (October 1977).

4. M. Nevière, D. Maystre and P. Vincent, "Determination of the Leaky Modes of a Corrugated Waveguide: Application to the Study of Anomalies," J. Optics (Paris), Vol. 8, pp. 231-242 (1977).

5. V. Shah and T. Tamir, "Simplified Grating Model for the Study of Absorption Anomalies," J. Opt. Soc. Amer., Vol. 69, p. 1473 (October 1979).

6. M. Nevière and P. Vincent, "Brewster Phenomena in a Lossy Waveguide Just Under the Cut-Off Thickness," J. Optics (Paris), Vol. 11, pp. 153-159 (1980).

7. V. Shah and T. Tamir, "Beam Shift Theory of Anomalous Absorption at Leaky-Wave Structures," J. Opt. Soc. Amer., Vol. 70, p. 1606 (December 1980).

8. A. Amittay, P.D. Einziger and T. Tamir, "Experimental Observation of Anomalous Electromagnetic Absorption in Thin-Layered Media," Appl. Phys. Letters, Vol. 38, pp. 754-756 (May 1981).

9. V. Shah and T. Tamir, "Anomalous Absorption by Multi-Layered Media," Optics Commun., Vol. 37, pp. 383-387 (June 1981).

10. L.M. Brekhovskikh, "Waves in Layered Media," (Academic Press, 1980).

11. R. Petit (Editor), "Electromagnetic Theory of Gratings," Topics in Current Physics, Vol. 22 (Springer-Verlag, 1980).

12. P.H. Berning and A.F. Turner, "Induced Transmission in Absorbing Films Applied to Band Pass Filter Design," J. Opt. Soc. Amer., Vol. 47, pp. 230-239 (March 1957).

13. H. Dupoisot, J. Morizet and P. Lostis, "Optical Interference Filters With Large Azimuthal and Spectral Widths," (in French), Appl. Optics, Vol. 13, pp. 1605-1609 (July 1974).

14. P.H. Lissberger, "Coatings with Induced Transmission," Appl. Optics, Vol. 20, pp. 95-104 (January 1981).

15. T. Tamir (Editor), "Integrated Optics," Topics in Applied Physics, Vol. 8 (Springer-Verlag, 1982), Chap. 3, p. 84.

16. K.C. Chang, V. Shah and T. Tamir, "Scattering and Guiding of Waves by Dielectric Gratings with Arbitrary Profiles," J. Opt. Soc. Amer., Vol. 70, pp. 804-813 (July 1980).

17. L.B. Felsen and N. Marcuvitz, "Radiation and Scattering of Waves," (Prentice-Hall, 1973) Chap. 4, p. 370.

18. T. Tamir and H.L. Bertoni, "Lateral Displacement of Optical Beams at Multilayered and Periodic Media," J. Opt. Soc. Amer., Vol. 31, pp. 1397-1413 (October 1971).

19. A. Saad, H. Bertoni and T. Tamir, "Beam Scattering by Non-Uniform Leaky-Wave Structures," Proc. IEEE, Vol. 62, pp. 1552-1561 (November 1974).

20. S.T. Peng, T. Tamir and H.L. Bertoni, "Theory of Periodic Dielectric Waveguides," IEEE Trans. Microwave Theory and Techniques, Vol. MTT-23, pp. 123-133 (January 1975).

21. R.S. Chu and T. Tamir, "Diffraction of Gaussian Beams by Periodically Modulated Media," J. Opt. Soc. Am., Vol. 66, pp. 220-226 (March 1976) and pp. 1438-1440 (December 1976); also, in Vol. 67, pp. 1555-1561 (November 1977).

22. T. Tamir and S.T. Peng, "Analysis and Design of Grating Couplers," Applied Physics, Vol. 14, pp. 235-254 (November 1977).

23. T. Tamir, "Guided-Wave Methods for Optical Configurations," Applied Physics, Vol. 25, pp. 201-210 (July 1981).

24. K. Ogusu, M. Miyagi and S. Nishida, "Leaky TE Modes on an Asymmetric Three-Layered Slab Waveguide," J. Opt. Soc. Amer., Vol. 70, pp. 48-53 (January 1980).

25. V. Shah and T. Tamir, "Leaky-Wave Approach to Induced Transmission", J. Opt. Soc. Amer., Vol. 71, p. 1574 (December 1981).

26. V. Shah and T. Tamir, "Anomalous Absorption Effects for Beams Incident on Lossy Layered Structures", 1982 Joint Intern. IEEE/APS and URSI Symp., Albuquerque, N.M. (May 1982).

27. V. Shah and T. Tamir, "Absorption and Lateral Displacement of Beams Incident on Multilayered Media", J. Opt. Soc. Amer., Vol. 73, pp. 37-44 (January 1983).

28. C.W. Hsue and T. Tamir, "Anomalous Scattering by Layered Media;" National Radio Sci. Meeting, Boulder, Colo. (January 1983).

29. T. Tamir and C.W. Hsue, "Leaky-Wave Interaction Effects for Beams Incident upon Thin-Film Structures", Tech. Dig. 4th Intern. Conf. Integrated Optics and Optical Fiber Commun., pp. 352-353, Tokyo, Japan (June 1983).

30. C.W. Hsue and T. Tamir, "Beam Scattering by a Layered Structure," Proc. 1983 URSI Intern. Symp. Electromagnetic Theory, pp. 191-194, Santiago de Compostela, Spain (August 1983).

31. C.W. Hsue and T. Tamir, "Lateral Displacement of Beams Refracted by Layered Media," J. Opt. Soc. Amer., Vol. 83, p. 1913; December 1983.

32. J.M. Zavada, J.J. Fasano, B.H. Li, M.J. Shiau and S.T. Peng, "Strong Absorption of Light by Multilayered Dielectric Metal Structures," Proc. SPIE Meeting, Arlington, VA (April 1983).

33. C.W. Hsue and T. Tamir, "Evolution of TE Surface and Leaky Waves in an Asymmetric Layered Configuration," J. Opt. Soc. Am., to be published in Sept. 1984.

34. C.W. Hsue and T. Tamir, "Lateral Beam Displacements in Transmitting Layered Structures," Optics Commun., Vol. 49, pp. 383-387; April 1984.

35. C.W. Hsue and T. Tamir, "Lateral Displacement and Distortion of Beams Incident Upon a Transmitting Layer Configuration," submitted to J. Opt. Soc. Amer.

36. C.W. Hsue and T. Tamir, "Power Flux of Gaussian Beams Incident on a Layered Structure," 1984 Meeting of the Opt. Soc. Amer., San Diego, CA; October 1984 (accepted for presentation).

C. NONUNIFORM OPEN DIELECTRIC WAVEGUIDES: TRANSITIONS AND TAPERS

Professor S.T. Peng

Unit EM4-3

1. OBJECTIVE(S)

A major problem in open dielectric waveguide systems that are under consideration for millimeter-wave integrated circuits is that unwanted scattering of waves can occur at any sharp discontinuity or junction. It is therefore essential that connections between circuit elements in such systems employ gradual transition structures in order to reduce the scattering losses. Some theoretical investigations have been made of such transition structures in open dielectric regions, but they have been quite limited due to the mathematical complexity in analyzing such transition structures in general. We propose to obtain analytical solutions for various types of transitions, tapers and junctions on open dielectric waveguides, by viewing these transition structures as sections of nonuniform waveguides. We intend to employ several different approximate procedures in these analyses, including a new one which was developed by us specifically for this program.

2. APPROACH

The transition structures to be studied will be viewed as nonuniform open dielectric waveguides, and mathematical techniques appropriate to that viewpoint will be employed. In particular, various approximations to be used will be examined critically and their limitations and usefulness assessed. To this end, specific models that are amenable to a rigorous analysis will be developed so that reliable and accurate results can be obtained as a standard against which approximate results can be compared and their validity judged. The program includes both the study of the validity and utility of the exact and approximate methods and their application to a number of specific transition structures for open dielectric waveguides.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

In the last report, we have presented an exact formulation of the wave propagation in a nonuniform periodic medium. More specifically, it is an almost periodic medium with a two-harmonic modulation, which can be considered as an extension of the well known canonical periodic medium with a sinusoidal modulation. Such a model is intended for the problem of tapering a periodic dielectric antenna in order to shape the radiating beam and to control the sidelobe level and distribution. Such tapering is usually designed by employing a locally uniform analysis. The approach investigated here should be capable of yielding an exact solution, against which the accuracy of the customary locally-uniform analysis can be compared in specific cases.

This new method, using a combination of a three-term recurrence relation in matrix form and the technique of noncommutative continued fractions, is employed to formulate the problem rigorously. Specifically, the new method offers the following advantages: (1) the analysis of the almost periodic medium is in a form analogous to that of the known Mathieu equation for the simpler canonical problem, (2) it applies to both commensurate and noncommensurate cases of the almost periodic medium, and (3) it yields a simple and effective algorithm for accurate numerical computations. Therefore, this method provides a general and unified approach to the solution of wave propagation in an almost periodic medium with a two-harmonic modulation.

While an exact formulation of the canonical problem of wave propagation in the almost periodic medium is of theoretical interest in its own right, the main purpose of such an undertaking is to prepare for the attack on a class of multilayer electromagnetic structures that contain periodic perturbations of different periods. As a continuing effort to gain better understanding of the wave phenomena taking place in an almost periodic structure, we have moved one important step forward during the past year. We have succeeded in our search for a mathematical model of an almost periodic structure, which is amenable to an exact formulation. Based on an exact dispersion relation, an extensive numerical analysis has been carried out to illustrate many interesting wave-propagation characteristics of almost periodic structures. A summary of the main results is given in what follows.

4. STATE OF THE ART AND PROGRESS DETAILS

A. Introduction

As shown in the previous Annual Report, the characteristic solutions of a doubly periodic medium are generally represented by a double Fourier series, accounting for the two different periods of the perturbation. In general, the number of transverse mode functions of a periodic medium is equal to the number of space harmonics that can be excited. Therefore, it is expected that the mode functions of the doubly periodic medium will be quadruply infinite in number. As is well known in the case of singly periodic structures,[1,2] the transverse mode functions, each of which consists of doubly infinite space harmonics, are coupled at a boundary discontinuity, resulting in two-dimensional (2D) coupling matrices describing the scattering of electromagnetic waves by a periodic structure. In contrast, in the present case of doubly periodic structures, the boundary conditions lead to three-dimensional (3D) matrices characterizing the coupling of space harmonics that are quadruply infinite in number. Thus, the boundary-value problem involving an almost periodic medium becomes considerably more difficult than the well-known singly periodic case; it requires a major effort to tackle the problem satisfactorily, if almost periodic structures are to be disigned for practical device applications.

More specifically, to treat the boundary-value problem of almost periodic structures, there exist two basic mathematical problems that demand a careful study:

(1) Systematic enumerations of the mode functions in such a way that will distinguish the relative importance of the modes.

(2) Development of an effective method for handling the 3D coupling matrices.

We are continually exploring these two mathematical problems for almost periodic structures in general. On the other hand, as a part of this research program, we have been making efforts to search for new models that are amenable to rigorous but simple formulations. We have succeeded in the latter effort. The new model we have developed will have a very significant theoretical impact, because it can provide very accurate results against which the accuracy of other approximate analyses can be evaluated. In addition, we present here some new and very interesting physical phenomena that may have very significant implications for the design of devices in the future.

B. Formulation

Consider a dielectric layer that is periodically corrugated on both boundary surfaces with two different periods, as shown in Fig. 1. Both periodic layers can be regarded as dielectric gratings consisting of uniform rectangular rods placed periodically in the x direction. The period and thickness of the upper grating (or grating A) are a and t_a and those of the lower grating (or grating B) are b and t_b, respectively. The corrugated structure has a dielectric constant ε_f, and the central portion, which can be taken as a uniform layer (film), has a thickness t_f. Such a corrugated structure is generally sandwiched between two half-space dielectric media. For simplicity, the upper and lower half-spaces will be referred to as the air and substrate regions, and their dielectric constants are denoted by ε_a and ε_s, respectively. Although Fig. 1 shows the special case of rectangular corrugations of the dielectric film, the theory to be presented below holds for dielectric gratings of arbitrary profile.

It is noticed that the structure under consideration contains two physically separated gratings of different periods. Furthermore, the guiding of waves along the doubly periodic structure may be viewed as a process of multiple reflections by the two gratings. Therefore, the formulation of the doubly periodic structure can be built on the knowledge of single periodic structures, as we shall do next.

In a doubly periodic structure, the electromagnetic fields everywhere consist of all the space harmonics due to the two different periods of the structure. In the uniform film region sandwiched between the two gratings, as shown in Fig. 1, the three components of the electromagnetic fields for the TE mode can be represented by the double Fourier series as:

$$E_y(x,z) = \Sigma\Sigma\, V_{mn}(z)\, \exp(-jk_{xmn}x) \tag{1}$$

$$H_x(x,z) = \Sigma\Sigma\, I_{mn}(z)\, \exp(-jk_{xmn}x) \tag{2}$$

$$H_z(x,z) = \Sigma\Sigma\, (k_{xmn}/\omega\mu)\, V_{mn}(z)\, \exp(-jk_{xmn}x) \tag{3}$$

with

$$k_{xmn} = k_{xoo} + 2m\pi/a + 2n\pi/b \tag{4}$$

Here, the summations are over m and n and both run from $-\infty$ to ∞, k_{xmn} is the propagation constant of the mn-th space harmonic in the x-direction, and V_{mn} and I_{mn} will be referred to as the voltage and current of the mn-th space harmonic, because they satisfy the transmission-line equations, for every set of integers m,n=0, ±1, ±2,...,

$$\frac{d}{dz} V_{mn} = -jk_{zmn} Z_{mn} I_{mn} \tag{5}$$

$$\frac{d}{dz} I_{mn} = -jk_{zmn} Y_{mn} V_{mn} \tag{6}$$

where k_{zmn} is the propagation constant in the z-direction, and Z_{mn} and Y_{mn} are the characteristic impedance and characteristic admittance of the transmission line, respectively. These quantities are related to k_{xmn} by:

$$k_{zmn} = (k_0^2 - k_{xmn}^2)^{1/2} \tag{7}$$

$$Z_{mn} = 1/Y_{mn} = \omega\mu_0/k_{zmn} \tag{8}$$

It is noted that expressions similar to those above may also be obtained for the TM mode, by duality of the Maxwell equations, but they are omitted here for simplicity.

In general, the electromagnetic fields in the film region consist of both forward and backward traveling waves in the z-direction. The voltage and current of the mn-th space harmonic can be obtained from the transmission-line equations, (5) and (6), as

$$V_{mn}(z) = f_{mn} \exp(-jk_{zmn}z) + g_{mn} \exp(jk_{zmn}z) \tag{9}$$

$$I_{mn}(z) = Y_{mn} [f_{mn} \exp(-jk_{zmn}z) - g_{mn} \exp(jk_{zmn}z)] \tag{10}$$

where f_{mn} and g_{mn} are the amplitudes of the forward and backward waves, respectively. Together with (1) and (2), (9) and (10) show that every space harmonic propagates in the uniform film region as an independent plane wave. However, these space harmonics are coupled by the two gratings, and the characterization of the couplings for the forward and backward wave amplitudes is the key step to the solution of the doubly periodic structure.

From the physical picture of wave processes explained above, we may now consider the scattering of each space harmonic of grating A by grating B and vice versa. Such a scattering problem has been extensively analyzed in the literature. In particular, under the JSEP, we have carried out a rigorous formulation of the single grating problem in the context of optical periodic couplers and millimeter-wave antennas.[1,2] Therefore, the previous results, including well-developed computer programs, are readily available for this new investigation.

It is noticed that between the two indices m and n in the expressions above, the first one, m, refers to the space harmonics of grating A and the second one, n, refers to those of grating B. Consider now the n-th space harmonic of grating B. Let f_{mn}, for m=0, ±1, ±2,..., be the set of amplitudes of the (incident) forward propagating space harmonics of grating A. Such a set of waves will be reflected by grating A, and the amplitudes of the (reflected) backward propagating space harmonics are g_{mn}, for m=0, ±1, ±2,.... For the present case, it has been shown[1] that the set of reflected harmonic amplitudes are related to the incident harmonic amplitudes by a reflection matrix as:

$$\underline{G}_n = \exp(-jK_n t_f)\, R_n\, \exp(-jK_n t_f)\, \underline{F}_n \tag{11}$$

where $\underline{F}_n$ and $\underline{G}_n$ are column vectors with f_{mn} and g_{mn} as their m-th elements, respectively, K_n is the diagonal matrix with k_{zmn} at the m-th diagonal entry, and R_n is the reflection matrix which is considered well established and can readily be evaluated numerically by our previously developed computer program for singly periodic structures. Such a reflection matrix has to be evaluated individually for every space harmonic of grating B, e.g., for every n=0, ±1, ±2,.... Thus, the complete set of harmonic amplitudes of backward-traveling waves is related to that of the forward-traveling ones by using the scattering characteristics of grating A.

Similarly, the electromagnetic waves are scattered by grating B. In this case, the backward-traveling waves are considered to be the incident waves and the forward-traveling waves become the reflected waves. Following the same procedure described above, for the m-th space harmonic of grating A, we can define a reflection matrix $S^{(m)}$, such that

$$\underline{F}^{(m)} = S^{(m)}\, \underline{G}^{(m)} \tag{12}$$

where $\underline{F}^{(m)}$ and $\underline{G}^{(m)}$ are column vectors with f_{mn} and g_{mn} as their n-th elements, respectively. Thus, we have obtained two independent sets of relations for the two sets of harmonic amplitudes f_{mn} and g_{mn}, for m, n=0, ±1, ±2,.... However, in (11) and (12), the two sets of column vectors ($\underline{F}_n$, $\underline{G}_n$), for n=0, ±1, ±2,..., and ($\underline{F}^{(m)}$, $\underline{G}^{(m)}$), for m=0, ±1, ±2,..., are different from each other, although they contain the same set of space harmonic amplitudes (f_{mn}, g_{mn}), for m, n=0, ±1, ±2,.... To convert $\underline{F}^{(m)}$ and $\underline{G}^{(m)}$, for m=0, ±1, ±2,..., into the other set, (12) can be rewritten explicitly in terms of the elements of the reflection matrix and the column vectors as:

$$f_{mn} = \sum_{i=-\infty}^{\infty} S_{ni}^{(m)} g_{mi} \tag{13}$$

for every m, n=0, ±1, ±2,.... Keeping the second index n fixed and collecting together all f_{mn} and g_{mn}, for m=0, ±1, ±2,..., the last equation can be written alternatively in the following form

$$\underline{F}_n = \sum_{i=-\infty}^{\infty} S_{ni}\, \underline{G}_i \tag{14}$$

where S_{ni} is a diagonal matrix with $S_{ni}^{(m)}$ as its element at the m-th diagonal entry. Now, (11) and (14) are two independent equations for the same two sets of vectors, $\underline{F}_n$ and $\underline{G}_n$. We may then eliminate $\underline{F}_n$ from these two sets of equations to yield

$$\underline{G}_n = \exp(-jK_n t_f)\, R_n \exp(-jK_n t_f) \sum_{i=-\infty}^{\infty} S_{ni}\, \underline{G}_i \tag{15}$$

for every integer n. Thus, (15) may be taken as a super system of homogeneous equations to determine the column vectors $\underline{G}_n$, n=0, ±1, ±2,.... The condition for the existence of a nontrivial solution is the vanishing of the coefficient determinant, e.g.,

$$\det[U\, \delta_{ni} - \exp(-jK_n t_f)\, R_n \exp(-jK_n t_f)\, S_{ni}] = 0 \tag{16}$$

where U is the unit matrix of infinite order and δ_{ni} is the Kronecker delta, which is equal to unity when n=i; otherwise, it is equal to zero. Equation (16) defines the dispersion relation for the doubly periodic structure.

C. Numerical Results

We have begun to carry out a numerical analysis of the guiding characteristics of the doubly periodic structure, based on the exact dispersion relation (16). We first investigate the case of bound waves for filter applications. For the structure shown in Fig. 1, and with the period a of grating A as a unit, we choose the following set of parameters: b=1.1a, for the period of grating B; $t_a=t_b=0.4a$, for the thicknesses of gratings A and B; $t_f=1.8a$, for the thickness of the film. In addition, the refractive indices of the materials are chosen to be: $n_a=n_s=1.0$ and $n_f=1.9$. For such a structure, we have examined the propagation characteristics in general, but we focus our attention on the stopband regions which are of primary interest for filter designs.

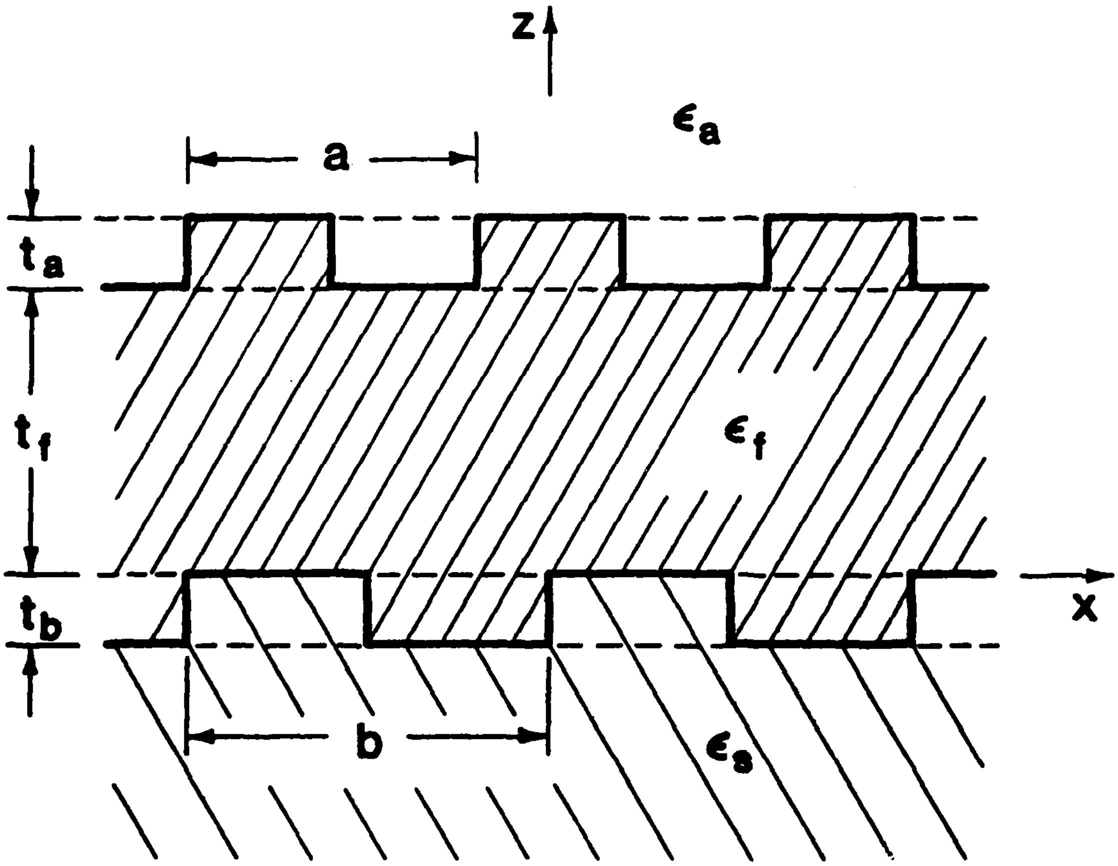


Fig. 1 Configuration of doubly periodic structure.

Before embarking on an elaborate analysis of the dispersion relation (16), it is very desirable to determine first an unperturbed dispersion curve by considering a uniform dielectric waveguide as an approximation for the almost periodic structure. The uniform dielectric waveguide is usually referred to as the unperturbed structure. Here, we take the average-thickness approximation by replacing each of the two periodic layers by a uniform one with a half thickness. The dispersion characteristics of the uniform waveguide are well known; for the present case, the dispersion curves of the lowest two TE modes are shown in Fig. 2, with the labels (1,0,0) and (2,0,0). Each one of these is regarded as the fundamental harmonic of the mode. In the presence of grating A, other harmonics with a phase shift of a multiple of $2\pi/a$ are excited; the dispersion curves for the $m=-1$ space harmonics for the two lowest modes are shown with the labels (1,-1,0) and (2,-1,0). Similarly, the dispersion curves for $n=-1$ space harmonics due to grating B are shifted by $2\pi/b$ and are labeled by (1,0,-1) and (2,0,-1). Among the three integers in each labeling, the first is the mode index, the second and the third are the harmonic indices due to gratings A and B, respectively. It is well known from the theory of mode coupling that the dispersion curves of a perturbed structure will follow closely those of the unperturbed structure, except in the vicinity of the intersection points of the unperturbed dispersion curves. Near the intersection points, the space harmonics are strongly coupled, resulting in complex dispersion roots that represent decaying waves due to total reflections, known as the Bragg Phenomenon.

It is noted that the structure parameters for the present analysis were chosen on the basis of these unperturbed dispersion curves. They were chosen such that the intersection of three unperturbed dispersion curves at the same point can occur, as shown in the boxed area marked by A. This means that in the case of doubly periodic structures three-mode coupling, which cannot be realized by singly periodic structures, can now be achieved. The three-mode coupling mechanism had been suggested as a means for the enhancement of mode coupling.[3] We have systematically investigated this potential three-mode coupling case, and the detailed stopband structures are shown as insets in Fig. 2. In Inset 1, we examine the two limiting cases, $t_a=0$ and $t_b=0$, corresponding to two singly periodic structures where only two-mode couplings can take place. As shown, we have two usual stopbands for the two limiting cases. Inset 2 shows the stopbands of the doubly periodic structure. Evidently, in the presence of the two gratings, all the modes are strongly coupled, so that the two stopbands occur at the same phase constant, as expected. Comparing the two insets, we observe that, contrary to expectation, one stopband is reduced in the case of the doubly periodic structure, while the other stays practically unchanged. More importantly, in the case of doubly periodic structures, in the region of overlap between the two stopbands, we have two different complex roots. Since complex roots must occur in pairs, this means that there exist four roots, with one extra or unexpected root, for the three-mode coupling case. Since our dispersion relation, (16), is an exact one, we may analyze the dispersion characteristics on a global or local basis. This enables us to explain such an apparent paradox that there exist four complex roots in the case of three-mode coupling. If we examine the overall space harmonics, it can be shown that for the case under consideration, it must be a four-mode interaction which is

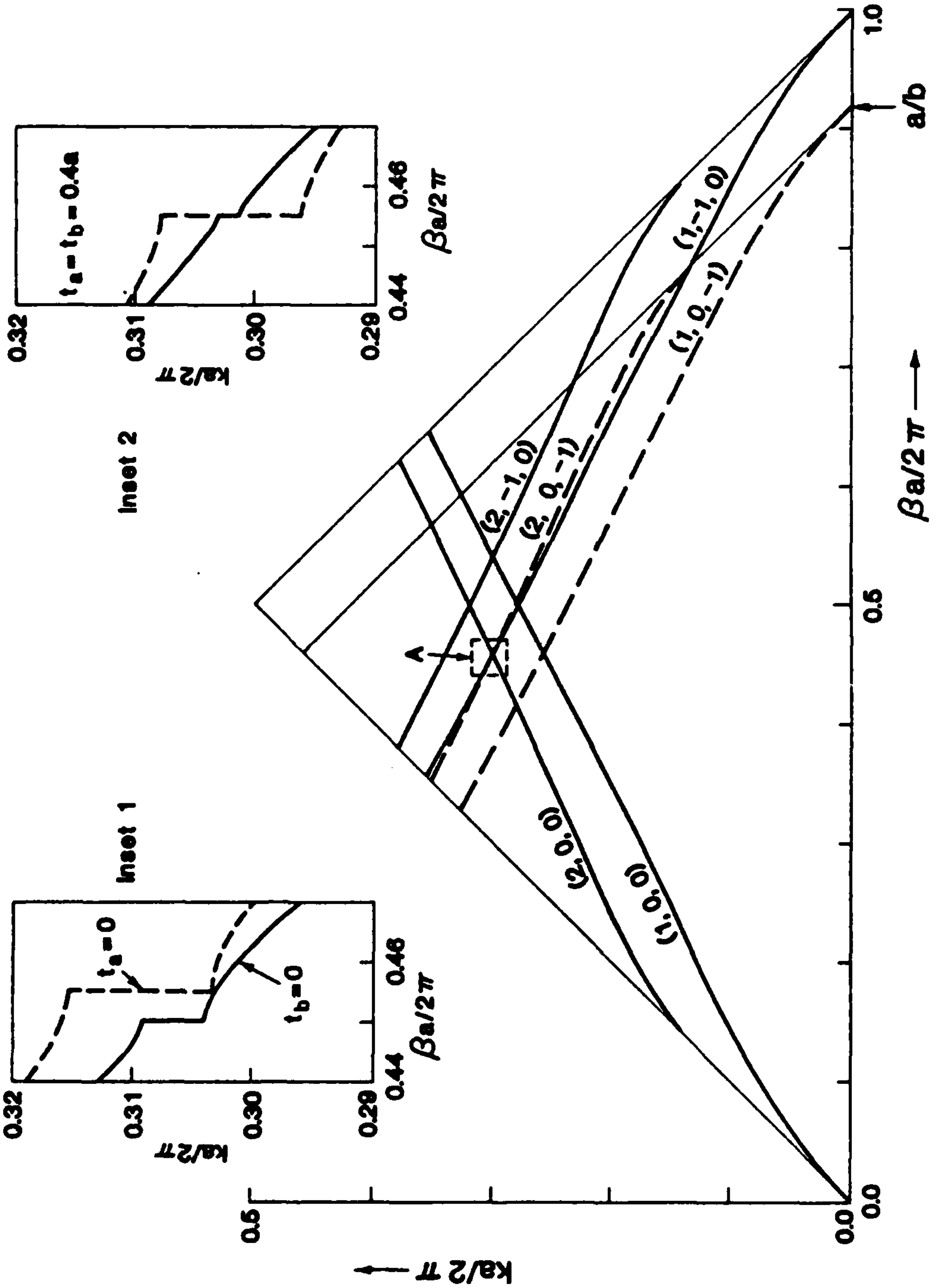


Fig. 2 Dispersion curves of doubly periodic structure.

automatically accounted for in our exact dispersion relation. We are pleased with the power of our approach that shows the unexpected results.

5. REFERENCES

1. S.T. Peng, T. Tamir, and H.L. Bertoni, "Theory of Optical Periodic Dielectric Waveguides," IEEE Trans. Microwave Theory Tech. Vol. MTT-23, p. 123, 1975.

2. F.K. Schwering and S.T. Peng, "Design of Dielectric Grating Antennas for Millimeter-Wave Applications," IEEE Trans. Microwave Theory Tech., Vol. MTT-31, pp. 199-209, 1983.

3. O.R. Asfar and A.H. Nayfeh, "Stopbands of the First-order Bragg Interaction in a Parallel-Plate Waveguide Having Multi-periodic Wall corrugations," IEEE Trans. Microwave Theory Tech., Vol. MTT-28, pp. 1187-1191, 1980.

D. COLLECTIVE FORMULATION OF WAVE PHENOMENA FOR GUIDING AND TRANSMISSION

Professor L.B. Felsen

Unit EM 4-4

1. OBJECTIVE(S)

High frequency propagation in, or transmission through, layered media usually requires synthesis in terms of a large number of basic wave processes. For the guiding or ducting problem, these wave processes are either normal (discrete and continuous) modes or ray-optical fields. For the transmission problem, the basic wave processes are traveling waves which undergo multiple internal reflection at the layer boundaries. Because descriptions by multiple propagation events are usually poorly convergent and do not provide physical interpretation in compact form, it is desirable to seek collective descriptions of multiple phenomena.

Thus, the objective of this fundamental study is the construction of a new theory of propagation, transmission and scattering that has broad implications for a general class of electromagnetic and other wave problems.

2. APPROACH

The technique for point source excitation problems has been to apply partial Poisson summation to a group of wave events. The Poisson formula expresses these events rigorously in terms of their Fourier transforms, and of truncation (remainder) terms. The Fourier transforms provide wave events of another type whereas the remainder terms yield collectively weighted wave events of the original type. The relative importance of the Fourier transformed or collectively weighted original wave types determines the nature of the final field representation. This novel and general procedure contains as a special case our previously developed hybrid ray-mode formulation. The general procedure is also to be explored within the framework of rigorous contour integral representations when these are available.

When the source is an aperture with tapered illumination which generates a well-collimated beam, simulation of the source field distribution by a local evanescent plane wave profile permits collective and direct tracking of the field via the theory of complex rays, without the need for an integration over the plane wave spectral components. For such excitations, this collective approach to the source problem is to be combined with that described above for the propagation or transmission problem.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contribution can be understood more clearly.

SECTION I: ELECTROMAGNETICS

The theory described briefly in the next section has been applied to a number of propagation, transmission and scattering problems to confirm its versatility and to test its accuracy. The accuracy tests have been performed by comparisons with solutions for canonical problems that can be generated by other means: numerical evaluation of rigorous series or integral representations, when available, or purely numerical solutions for (non-separable) problems that do not yield analytic solutions in explicit form. These applications are summarized below. All of the studies performed so far are for two-dimensional configurations in order to eliminate complexities in detail that are not essential in establishing the basic validity of the collective approach and of the ray-mode equivalent. A comprehensive survey is contained in an invited paper to be published in the IEEE Transactions on Antennas and Propagation.[1]

1) Time-Harmonic Fields

(a) Singularity-free field tracking

One class of problems deals with repairing the deficiencies in an asymptotic ray theory (ART) analysis of high-frequency propagation in a highly overmoded waveguide that has an inhomogeneous refractive index profile in the transverse direction x. If the refractive index decreases monotonically from the perfectly reflecting top boundary to the bottom boundary, which separates the waveguide from an exterior semi-infinite medium with (lower) constant refractive index, one encounters the following categories of rays originating at the source: a) rays which are continuously refracted toward the top without encountering the bottom; b) rays which encounter top and bottom but are totally reflected at the bottom; c) rays which encounter top and bottom but are refracted into the exterior. ART fails (it predicts infinite fields) in the following transition regions: near the caustics formed by the surface guided rays in category a); near the bottom-glancing ray that separates categories a) and b); near the critically incident ray that separates categories b) and c). By the ray-mode equivalent, spectral intervals surrounding these transitional rays are filled with modes. The theory has been developed and implemented numerically for a model waveguide with exponentially varying index profile.[2]

(b) Waveguides with longitudinal variation

When the refractive index and (or) the waveguide boundaries change along the guiding direction z, the resulting generally non-separable boundary value problem can be attacked by coupled mode theory provided that one invokes discretizing approximations (by inserting a false boundary in the exterior medium) to eliminate the continuous mode spectrum. The modes in question are those pertaining to the local z-independent environment. To reduce the intermode coupling, it would be useful to search for a more general class of local modes that adapts better to the changing environment than those in the plane parallel model. Such "more adaptable" modes are the "adiabatic" modes which, to the lowest order in the non-separability parameter, propagate without coupling to other adiabatic modes. Thus, they form an uncoupled system in waveguides with weak longitudinal variation. By two different approaches, one based on collective treatment of ray spectra, the other on scaling of mode spectra, we have developed a new theory for tracking these modes, with inclusion of the continuous spectrum, along the waveguide and also through the cutoff region if the configuration is such that an originally trapped adiabatic mode gets converted into a radiating (leaky) mode. These two approaches, which can be shown to yield the same asymptotic solution, have been published.[3,4] Efforts during the past year have been spent on refinements and numerical implementation.

(c) Complex ray modeling of distributed aperture fields

When the source of radiation is distributed over a large aperture instead of being localized, conventional ray tracing techniques become cumbersome and even invalid when the aperture field is tapered so as to generate a collimated beam. To circumvent the need for integration over the large aperture surface to pass from the aperture to the far zone, we have proposed to simulate the aperture field by Gaussian-like distributions that can be generated by a source point in complex coordinate space or, more generally, by an initial distribution of complex rays. The resulting fields can then be traced from the analytically continued aperture to the far zone without integration over the aperture surface. Work on this complex ray simulation of various taper profiles has progressed and has been implemented for a class of profiles ranging from Gaussian to rectangular.[5] For the Gaussian-like fields generated by a complex source point, the tracing of the beam from the aperture through a cylindrical shell radome to the far zone has already been accomplished.

2) Transient Fields

(a) Hybrid wavefront-resonance formulation of transient scattering

By extending the concept of rays, modes and their equivalents into the transient domain, we have developed a new theory of transient scattering that combines wavefronts (rays) and complex resonances (modes) in a self-consistent framework for efficient analysis of a target response at all observation times. On the prototype problem of impulse scattering by a circular cylinder, we have shown how the cumulative effect of wavefront arrivals generates the complex resonances of the singularity expansion method (SEM) and, conversely, how the cumulative effect of resonances generates wavefronts. This new theory has been published,[6,7] and it is now being extended to strip obstacles.

(b) Spectral theory of transient fields

Progress has been made on reformulating transient propagation and scattering by inverting the conventional procedure, which first solves for a source-excited field in terms of a wavenumber spectral integral in the frequency domain and thereafter, by Fourier inversion, passes to the time domain. Certain advantages accrue by performing the frequency integration before the integration over the wavenumber spectrum. We have endeavored to generalize this approach, introduced by Chapman[8] for seismological problems, by viewing it within the context of weakly dispersive propagation processes. Our generalization of the Chapman procedure, which is constrained to real frequencies and spectral wavenumbers, involves analytic functions in the complex frequency and spectral wavenumber planes and therefore avoids Chapman's use of the theory of distributions and of Hilbert transforms. The basic ideas have been presented in a short publication[9] and are being developed further.

4. STATE OF THE ART AND PROGRESS DETAILS

A. Introduction

Many electromagnetic propagation environments, whether natural or man-made, are so complicated that direct solution of the field equations to determine signal characteristics is beyond the scope of present analytical and computational capabilities. A common procedure has been to decompose the complicated, intractable propagation process into a sequence of simpler tractable events, and to synthesize the original problem by superposition of, and interaction among, the simpler constituents. Examples are provided by propagation in tropospheric layers, in the earth-ionosphere waveguide, along the surface of the earth, in optical fibers, etc., where the presence of transverse medium inhomogeneities or boundaries causes multiple reflections or ducting of the source-excited radiation. Here, the simpler problem would be the radiated field in the absence of boundaries, with the influence of the latter accounted for by multipath effects. In such a guiding environment, the field may alternatively be expressed in terms of discrete and continuous guided modes, but <u>many</u> of these are required at high frequencies when the duct width is large compared to the local wavelength, and "non-canonical" environments may introduce mode coupling. A somewhat different viewpoint prevails for propagation through, rather than along, stratified media as, for example, in vertical ionospheric sounding or in propagation through layered dielectric radome covers for aperture antennas. Here, guiding effects are generally unimportant but emphasis is placed on multiple reflections due to successive interfaces or medium gradients along the transmission path. Moreover, with respect to the excitation, different propagation phenomena are relevant when the incident field is due to a localized source and therefore has a broad radiation pattern, or due to a distributed aperture that generates a well-collimated beam.

For the class of problems alluded to above, superposition (by multiple reflection) of individually tractable propagation events constitutes one of the principal tools of analysis. At high frequencies, these events may be approximated as ray fields which undergo reflection, refraction and(or) diffraction on their path from a source at S to an observer at P. While ray theory provides a fundamental view of the propagation process by tracking local plane wave fields emanating at the source, such tracking becomes cumbersome when many ray paths exist between S and P. It would therefore be desirable to deal with multipath effects <u>collectively</u>. In guided propagation <u>along</u> a refracting channel, rays may form caustics (convergence or focusing zones of enhanced field strength) where simple ray theory fails. When these caustics are sufficiently distinct, one may correct ray theory by uniform asymptotic transition functions, but situations arise for rays with many reflections where an accumulation of caustics makes such corrections impractical and even impossible. Here, again, a collective alternative to multiple ray reflections is desirable or actually necessary. In transmission <u>through</u> a layered medium, an attractive alternative to tracking individual internal interface reflections and refractions would be to deal with these collectively by defining a "local slab" transmission or reflection coefficient.

The modal approach to ducted propagation is beset with similar difficulties when the required number of modes is large. It would be useful to express the interference properties of clusters of modes collectively in terms of simpler events. Moreover, asymptotic mode fields in complicated refracting media possess caustics near which that simple description fails. Collective treatment of such mode clusters could alleviate this difficulty.

The above considerations pertain to excitation by a localized source with a broad radiation pattern. Many applications deal, however, with distributed aperture excitations that generate a well-collimated beam field. These latter excitations can be synthesized in terms of the point source fields either by simulated discrete superposition or by continuous integration over the aperture domain. Alternatively, the source field could be decomposed into a plane wave spectrum and the observed field recovered by spectral synthesis of the plane waves after they have traversed the propagation environment. Each of these approaches requires an additional superposition or integration, which one would like to avoid if the excitation is of the beam type. Phrased in another way, one would like to be able to deal <u>collectively</u> with the distributed source problem for focused aperture fields.

The preceding discussion has made evident the importance of collective treatment of mode or ray fields when many of these are required to synthesize the signal in a particular transmission or guided propagation channel, or when failures in approximate mode or ray theory make these descriptions inapplicable. Substantial progress in this direction has been made through the new hybrid ray-mode theory developed by us. It has been shown how clusters of ray fields excited by a localized source can <u>rigorously</u> be converted into clusters of guided mode fields plus a (usually small) remainder, and vice versa. The theory has been applied to a series of "canonical" problems involving guided electromagnetic propagation along concave surfaces (here, the guiding mechanism is provided by "whispering gallery" effects), in tropospheric ducts, in plane parallel homogeneously filled waveguides, and in graded index waveguides. The theory has also had impact on other fields such as underwater acoustic propagation and, with generalization to time-dependent signals, the modeling of seismic events. Concern in these applications has been with the greater computational efficiency of the hybrid formulation, with the avoidance of singular regions in ray fields or mode fields by filling these regions with mode fields or ray fields, respectively, and also with the penetrating physical insight of the propagation mechanism provided by the hybrid method. These accomplishments have been documented in a comprehensive series of publications.[10-22]

With respect to the source problem, it has been shown previously that well-collimated fields excited by tapered aperture illuminations can be analyzed effectively by use of complex rays and evanescent waves, without the need for aperture integrations or plane wave spectral synthesis. The present program deals in a <u>general</u> combined fashion with the <u>collective</u> description of a group of propagation events as well as a class of focused aperture excitations. The former was motivated by the hybrid ray-mode formulation and contains this formulation as a special case, whereas the latter is based on our previously performed evanescent wave and complex ray studies. We believe that application of the collective point of view to wave phenomena as well as distributed sources can make substantial impact on complicated propagation problems. The analytical setting for the collective wave treatment is provided, in general, by the Poisson summation technique, which is described briefly below. Further details may be found in reference 1.

B. Foundation of the Collective Treatment

The collective method provides an inherently rigorous scheme for combining ray fields and mode fields (with inclusion of a remainder, when necessary) in well-defined proportions. For laterally homogeneous but vertically stratified media, the method can be derived either from an initial plane wave spectral representation of the Green's function or by application of partial Poisson summation. Via the former route, one decomposes the integrand in the plane wave spectral formulation in various ways so as to generate via contour deformation generalized ray integrals, or normal and leaky mode contributions (including continuous spectra). When only some of the ray integral or modal contributions are retained, the effect of the omitted contributions can be expressed in terms of mode and ray fields, respectively, plus remainder terms. The necessary manipulations in the complex wavenumber plane to achieve this hybrid representation have been illustrated in the various publications mentioned earlier. Alternatively, one may proceed by applying Poisson summation to a selected group of ray or mode fields, without starting first from an integral representation of the Green's function. Because the Poisson summation route is generally applicable to determine the collective behavior of any sequence of terms, and also because it highlights the physical mechanisms, it is discussed briefly in *the following*.

The Poisson identity for a group of terms f_q can be written as:

$$\sum_{q=Q_1}^{Q_2} f_q = \frac{1}{2}\, f_{Q_1} + \frac{1}{2}\, f_{Q_2} + \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \int_{2\pi Q_1}^{2\pi Q_2} f(\tau) e^{ip\tau} d\tau \qquad (1)$$

In (1), $f(\tau)$ is obtained from f_q on replacing the discrete index q by $(\tau/2\pi)$, with τ representing a continuous variable. Let us assume that each f_q may possess in general a slowly varying amplitude A_q and a rapidly varying phase ψ_q so that

$$f_q = A_q \exp(i\, \psi_q) \qquad (2)$$

Alternatively, f_q could be defined in terms of an integral (for example, over a spectrum of plane waves), with the index q occurring in the integrand. In that event, that integration remains in the corresponding transition to $f(\tau)$ on the right-hand side of (1). It should be noted that f_q in wave problems also depends on other parameters, for example, source and observation points and medium parameters, and that the analytic behavior of f_q and $f(\tau)$ may be strongly affected by these. Evidently, the Poisson sum formula expresses the events f_q collectively in terms of their Fourier transforms with respect to the "smeared out" summation index, plus truncation effects depicted by (1/2) of the contributions from the first and last element.

To treat the τ-integral in (1), it is convenient to perform an asymptotic evaluation by the method of saddle points. With $f(\tau)$ expressed as in (2), contributions will arise from possible saddle points and from the endpoints. Details are presented in reference 1, with the result:

$$\sum_{q=Q_1}^{Q_2} f_q \sim \sum_p F_p + \frac{1}{2}\,(1 + \Delta_{Q_2})\, f_{Q_2} + \frac{1}{2}\,(1 - \Delta_{Q_1})\, f_{Q_1} \qquad (3)$$

where

$$\Delta_{Q_{1,2}} = i \cot(\pi\, d\psi/d\tau)_{\tau_{1,2}} \qquad (4)$$

and F_p represents the contribution from relevant saddle points (isolated or in clusters) in the Poisson integral while $\Delta_{Q_{1,2}}$ terms arise from endpoints. Each F_p is a collective alternative of f_q in that they are Fourier transforms of one another; for example, if f_q belongs to a family of rays, F_p belongs to a family of modes, and vice versa. On the other hand, the endpoints of the Poisson integral yield a collective description of the set $\{f_q\}$ in terms of the first and last elements, appropriately weighted, of the <u>same</u> set. Thus, if $\{f_q\}$ is a set of rays, then the endpoint contributions express its truncation as two equivalent <u>collective rays</u>. Similarly, a group of modes gives rise, due to truncation, to two equivalent <u>collective modes</u>. To the leading asymptotic order, the saddle point contributions dominate over the endpoints. Therefore, if relevant saddle points exist, the collective behavior of the set $\{f_q\}$ is predominantly characterized by the alternative set $\{F_p\}$, with the collective elements $f_{Q_{1,2}}$ providing some fine tuning.

When saddle points or singularities approach the endpoints $\tau_{1,2}$ of the integration interval, the simple formula in (3) must be modified. Evidently, such a circumstance defines a transition region, where one of the elements F_p may escape from, or enter, the group. Appropriate uniform functions (Fresnel integrals) ensure the smooth behavior of the total field as this transition takes place. When a saddle point coincides with an endpoint, the asymptotic result furnishes one-half of the saddle point contribution and there is then no endpoint contribution.

The results described above have a cogent physical interpretation which can be illustrated on a simple example. Consider a single homogeneous layer wherein a source S transmits a field to an observer at P. We concentrate on wave phenomena that leave the source in the upward direction and arrive at P from below. Out of the complete set of direct and multiple reflected rays f_q, we have retained only the direct ray and the first two reflected rays with angles $\theta_{1,2,3}$ at S; i.e., q = 1,2,3 (Figure 1(a)). These rays occupy the vertically shaded angular interval. The normal modes in the layer can be decomposed into upgoing and downgoing plane waves at the characteristic mode angles $\bar{\theta}_p$. In Fig. 1(b), it is assumed that only two adjacent modes, with characteristic angles $\bar{\theta}_1$ and $\bar{\theta}_2$, fall within the angular interval $\theta_3 < \bar{\theta}_p < \theta_1$; i.e., $\tau_{1,2}$ are the only saddle points on the integration interval in (1). Since the "mode volume" between $\bar{\theta}_1$ and $\bar{\theta}_2$ does not fill the initial "ray volume" between θ_1 and θ_3, there are intervals empty of modes (Figure 1(c)). Their effect is accounted for by the collective ray portions $(1/2)\Delta_{Q_{1,2}} f_{Q_{1,2}}$ in (3). If the mode volume fills the <u>entire</u> ray volume, then $\Delta_{Q_{1,2}}$ disappears and the first and last modes (here the only modes) contribute with half strength. The resulting reduction of (3) then exhibits <u>complete symmetry</u> between the group of rays and the group of modes, with the first and last element of each halved to account for truncation. Analogous considerations apply when an initial group of modes is to be converted into rays.

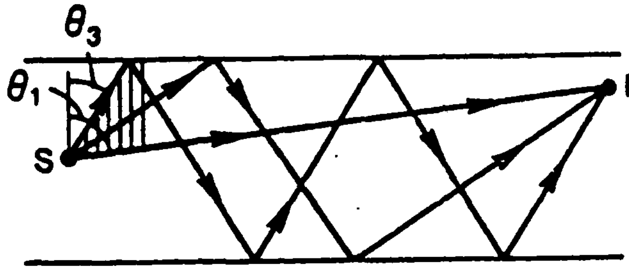
Thus, any group of rays can be converted into a well-defined group of modes and vice versa, with inclusion of collective rays or modes, when needed. This is the essence of the hybrid approach (see Figure 2). The asymptotic considerations above clarify the physical mechanism, but the formula in (1) is exact and can be used directly for numerical evaluation.

C. Recent Progress

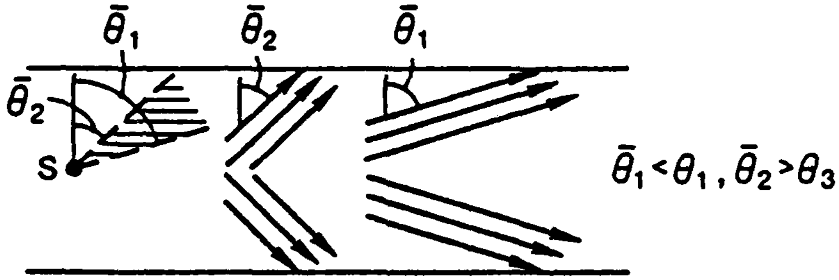
1) Repairing deficiencies of Asymptotic Ray Theory (ART)

An inhomogeneous waveguide with monotonically decreasing refractive index can trap rays and modes near the maximum index surface. The ray fields form caustics which may pile up at long distances from the source. The hybrid ray-mode approach can be employed to avoid inclusion of uniform asymptotic corrections of ray theory as the observer approaches the caustics. When the vertically inhomogeneous profile is terminated at $x = b$ by a semi-infinite medium with constant refractive index $n_2 < n(b)$, there occur new transition regions in the vicinity of the glancing ray and the reflected ray excited by a ray incident at the critical angle (Fig. 3).

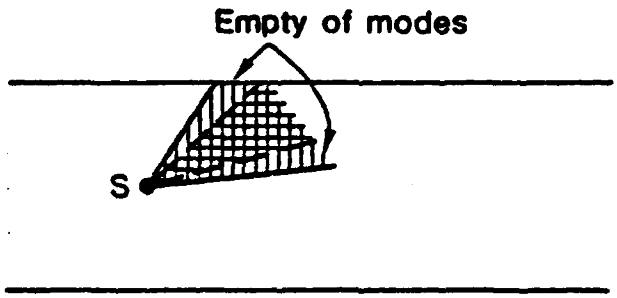
The failure of ray theory in these transitional domains can be corrected by replacing the angular spectra surrounding these rays by a bundle of modes plus collective remainders. These remainders can be phrased either as collective rays or collective modes (see Sec. B). For an application in seismology (Fig. 4), we have computed numerical results by the <u>hybrid</u> method and compared with direct numerical integration of the exact integrals in the formulation, which serve as a reference solution for comparison. The results are summarized in



(a) group of rays with departure angles θ_q



(b) corresponding group of modes with characteristic angles $\bar{\theta}_p$



(c) composite domain, with empty regions

Fig. 1. Ray-mode equivalence in a homogeneous layer. Three rays in the depicted category (upgoing at S and at P) are to be converted into modes. At a fixed frequency, the characteristic mode angles are constant but the ray-mode combination is range-dependent since the ray paths change with relative positions of S and P. In the time-dependent regime, the hybrid combination changes even for fixed S and P (fixed ray angles) since the characteristic mode angles now vary with frequency.

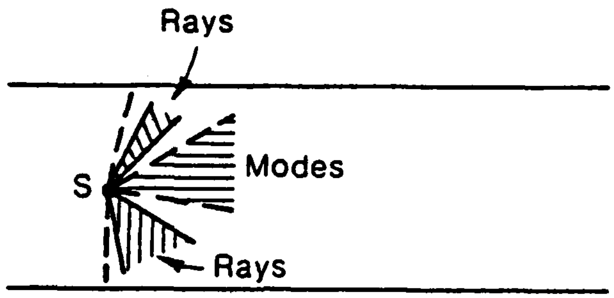


Fig. 2 Schematic representation of hybrid scheme. Alternate angular intervals at the source are filled with rays (vertical shading) or modes (horizontal shading). The effect of empty intervals (blank) is accounted for by collective rays or modes.

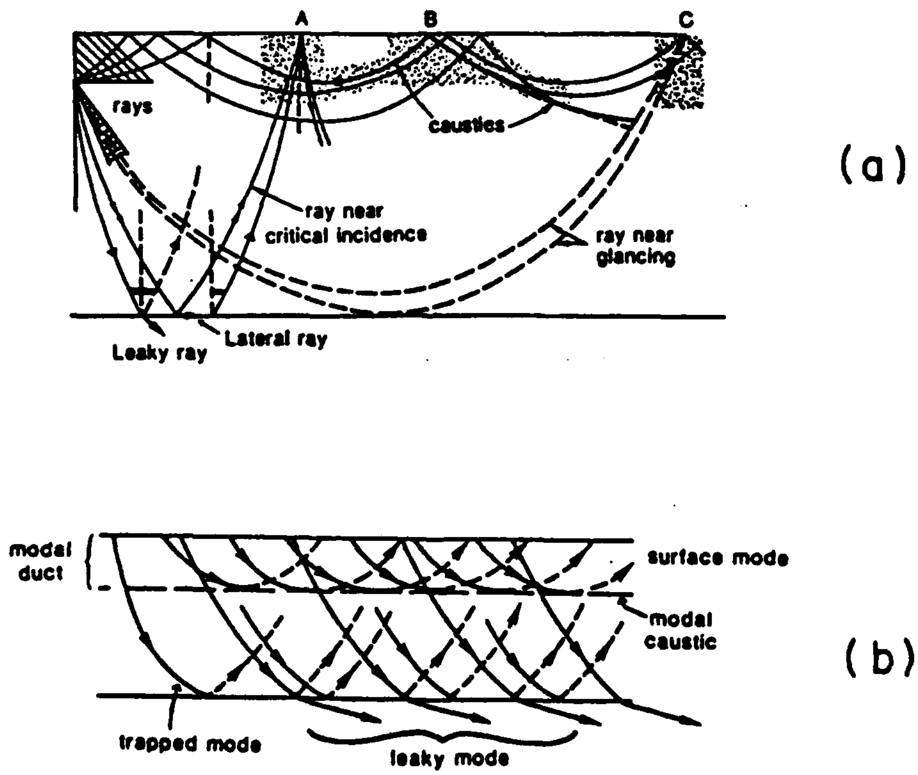


Fig. 3 Rays and modes in a waveguide with inhomogeneous refractive index that decreases monotonically with depth.
a) ray trajectories. Asymptotic ray theory fails in transition regions (dotted). The corresponding angular intervals (shaded) can be filled with modes in a hybrid representation.
b) modal ray congruences.

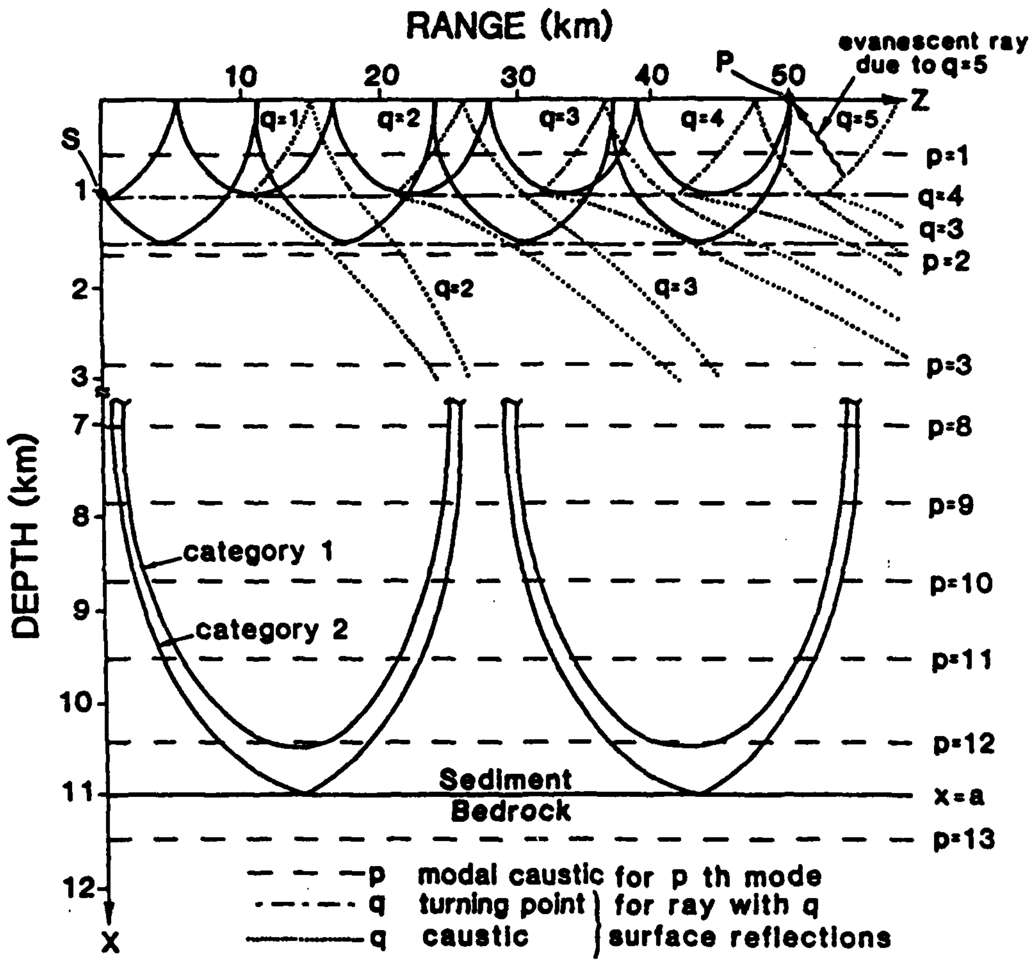


Fig. 4 Surface-ducted and bottom-glancing ray transition regions in a seismic model involving an inhomogeneous sediment layer above a homogeneous semi-infinite bedrock bottom at x = a = 11 km. A seismic SH source at a depth of 1 km operates at a frequency of f = 1.644 Hz. The shear velocity profile (reciprocal of refractive index) in the sediment is $v_s(x) = v_s(0)\exp(x/d)$, with $v_s(0) = 1.5$ km/sec, d = 11.0 km; thus, $v_s(a) = 3.2$ km/sec. The shear velocity in the bottom is $v_{sb} = 4.08$ km/sec.

Table I for the caustic forming rays and in Table II for the bottom glancing ray. For the indicated parameter range in Table I, the observer is reached by rays with up to four ($q = 4$) surface reflections; rays with $q \geq 5$ reach the observer by evanescent tunneling through the dark side of their caustics, and some of these rays may be significant. Rays with $q = 3,4$, lie so close to the surface that isolated caustic corrections for them are inapplicable. Consequently, ray fields with $q \geq 3$ must be treated in an alternative manner. Here, these fields are first computed by numerical integration (NI) of the exact ray integrals. This reference solution is then compared with results obtained by replacing these ray fields with modes plus uniformized collective mode remainders $(HY)_M$ and with modes plus uniformized collective ray remainders $(HY)_R$. Moreover, to illustrate the failure of simple asymptotic ray theory (ART), the ART field for ray $q = 4$ is compared with the exact value from numerical integration of its ray integral. It is seen that the $(HY)_M$ and $(HY)_R$ solutions agree well with the reference (NI) solution, and it should be noted that the computation time for the former is only 15% - 25% of that required for NI.

For observation points in region C of Fig. 3, transition phenomena arise due to two rays, one of which has just grazed and the other just missed the bottom. Therefore, these rays are treated together for description of the glancing ray transition. The numerical results are shown in Table II. For the chosen parameters, it was found that the critical ray transition region A and the glancing ray transition region C are in close proximity. Therefore, each cannot really be treated in isolation from the other. This circumstance affects especially the $(HY)_R$ equivalent, which is found to require a wider mode bundle then the $(HY)_M$ equivalent. For this reason, no data are shown for $(HY)_R$. Despite these limitations, the comparisons reveal again the adequacy of the modal equivalent, with the same gain in numerical efficiency noted previously. The comparison with ART is far less satisfactory.

2) <u>Range-dependent waveguides</u>

Substantial effort has been devoted to open waveguides with (non-separable) longitudinal variation. Here, the conventional approach has been via coupled mode theory. In that theory, one generally ignores coupling to the continuous radiation spectrum, although efforts have been made to discretize the spectrum by inserting artificial boundaries, including some loss, and then attempting to have the complex modes of this huge multimode waveguide describe the relevant wave process in terms of discrete spectra only. When, as in many physical problems, the longitudinal dependence is sufficiently slow over an interval of length equal to the local wavelength, simpler approximations involving adiabatic techniques can be applied. An adiabatic mode alters its character by smoothly adapting to the changing environment without, to the lowest order of approximation, coupling to other adiabatic modes. However, this simple description fails in transition regions where (in a narrowing waveguide, for example) an initially trapped mode encounters cutoff and is transformed into a radiating mode. The failure of adiabatic mode theory in the transition region has hampered its application to this important class of events. We have addressed the transition problem by two methods motivated by the collective approach to guided mode type

Table I Surface Ducted Ray Fields. Comparisons of asymptotic ray theory (ART), modal equivalents $(HY)_{M,R}$ (2-3 modes) for two caustic-forming rays, and exact results (NI). Field values are decomposed into real (Re) and imaginary (Im) parts.

	z = 50.0 km		z = 52.0 km		z = 53.0 km	
	Re	Im	Re	Im	Re	Im
$(NI)_4$	-0.352	-0.284	-0.270	-0.315	-0.230	-0.334
$(ART)_4$	-0.585	-0.309	-0.423	-0.374	-0.365	-0.392
NI	-0.217	-1.902	1.232	-1.790	1.838	-1.452
$(HY)_M$	-0.220	-1.927	1.250	-1.817	1.793	-1.477
$(HY)_R$	-0.213	-1.910	1.253	-1.826	1.854	-1.468

Table II Bottom Glancing Ray Fields. Comparisons of asymptotic ray theory (ART), modal equivalent $(HY)_M$ (6 modes) and exact results (NI) for the two transitional rays. Exact range for glancing ray is 61.2 km.

	z = 58.0 km		z = 60.0 km		z = 61.0 km	
	Re	Im	Re	Im	Re	Im
NI	-0.298	0.150	-0.172	-0.073	0.095	0.096
ART	-0.296	-0.059	-0.158	-0.311	0.082	0.353
$(HY)_M$	-0.294	0.144	-0.170	-0.088	0.098	0.119

and ray type events. The collective treatment in this general environment is based on the definition of an adiabatic invariant that ensures that the essential features of the collective process are retained intact.

The first method[4] is based on a plane wave spectral representation of the waveguide Green's function, scaled via a modal invariant so that the local plane waves synthesizing a local (adiabatic) mode adjust their characteristic propagation angles or wavenumbers to satisfy the local transverse resonance condition for the modal eigenvalue. This new concept allows the coordinate separable spectral integral for a plane parallel waveguide to be generalized to weakly non-separable (for example, slowly tapered) waveguide configurations. The resulting spectral integral contains all of the relevant propagation phenomena, including the transition of an adiabatic mode through cutoff. Principal effort has been expended on refining the spectral integral, and on developing a numerical routine for computing it for the special case of a wedge-shaped dielectric waveguide. These studies are now approaching completion.

The second method[3] proceeds from the recognition that ray fields in a longitudinally changing waveguide do not have to undergo spectral scaling since they describe ab initio a local propagation process. Ray fields can be constructed without constraints of separability, provided only that conditions change sufficiently slowly to validate local plane wave propagation along each trajectory defined by the ray equations. The ability to define local modes in a non-separable configuration can be tied to the ability to treat multiple reflected ray fields collectively by Poisson summation (Sec. 4B, Eq. (1)). This has been carried through in detail for a wedge-shaped dielectric waveguide. By a significant generalization, the previous analysis[3], based on a representation of the total source excited field, has now been adapted to accommodate an individual mode, which we call an intrinsic mode.[23] The intrinsic mode is given by a spectral integral which describes its progress uniformly from the trapped region through cutoff into the radiating regime. Below cutoff, the integral reduces to the local trapped normal mode; most of its field is here confined to the waveguide, with evanescent decay outside. As the local mode approaches cutoff, more of the field gets squeezed into the exterior, and well beyond cutoff, essentially all of the field forms an exterior beam. Major effort has been extended during the past year to develop an efficient algorithm for computing the intrinsic mode.[24] Preliminary results for the lowest mode in an ocean acoustic example are shown in Fig. 5. The results are compared there with numerical computations obtained elsewhere by the parabolic equation algorithm and by an "augmented adiabatic mode" algorithm. Because the spectral theory leading to the intrinsic mode is inherently exact for this example, we believe that whatever discrepancies exist are due to the other methods. It is to be emphasized here that for small taper angles, the intrinsic modes are uncoupled although the dielectric wedge waveguide is non-separable.

We regard the intrinsic mode scheme as a basic contribution to treating propagation in tapered open waveguides. The theory is now being explored further with a view toward generalization to other taper contours. For stronger tapers, the intrinsic modes will have to be coupled, but such coupling is expected to be much weaker than that between the conventional local plane parallel modes.

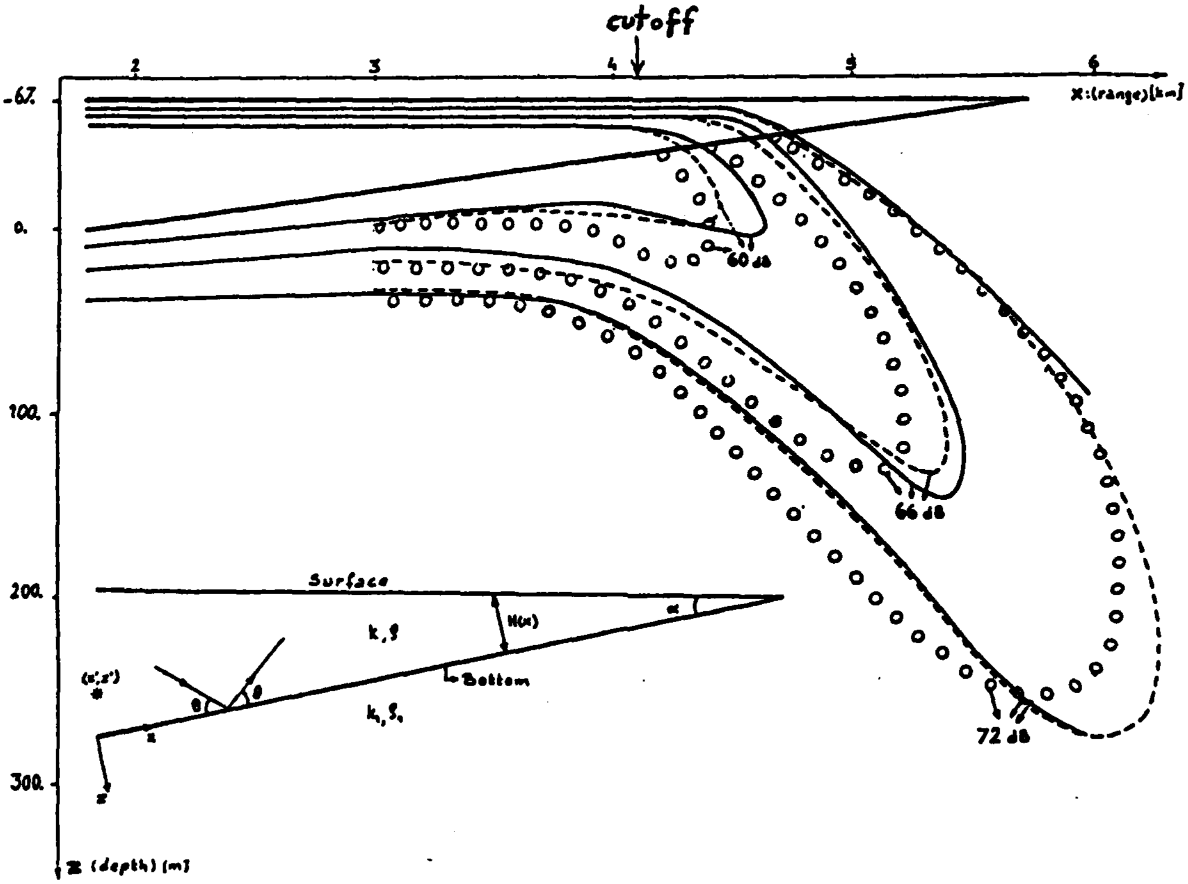


Fig. 5 Wedge-shaped ocean geometry.

5. REFERENCES

1. L.B. Felsen, "Progressing and Oscillatory Waves for Hybrid Synthesis of Source Excited Propagation and Diffraction," Invited Paper, to be published in IEEE Trans. on Antennas and Propagation.

2. E. Niver, A. Kamel and L.B. Felsen, "Modes to Replace Transitional Asymptotic Ray Fields in a Vertically Inhomogeneous Earth Model," to be published in Geophys. Journal of the Roy. Astron. Soc.

3. J.M. Arnold and L.B. Felsen, "Rays and Local Modes in a Wedge-Shaped Ocean," J. Acoust. Soc. Am., Vol. No. 73, No. 4, April, 1983 (1105-1119).

4. A. Kamel and L.B. Felsen, "Spectral Theory of Sound Propagation in an Ocean Channel with Weakly Sloping Bottom," J. Acoust. Soc. Am., Vol. 73, No. 4, April, 1983 (1120-1130).

5. G. Ghione, I. Montrosset and L.B. Felsen, "Complex Ray Analysis of Radiation from Large Apertures with Tapered Illumination," to be published in IEEE Trans. on Antennas and Propag.

6. E. Heyman and L.B. Felsen, "Traveling Wave and Oscillatory Formulations of Scattering Problems," accepted for publication in Handbook on Acoustic, Electromagnetic and Elastic Wave Scattering - Theory and Experiment, 1983.

7. E. Heyman and L.B. Felsen, "Creeping Waves and Resonances in Transient Scattering by Smooth Convex Objects," IEEE Trans. on Antennas and Propag., AP-31 (1983), pp. 426-437.

8. C.H. Chapman, "A New Method for Computing Synthetic Seismograms," Geophys. J. Roy. Astron. Soc., Vol. 54, pp. 481-518, (1978).

9. E. Heyman and L.B. Felsen, "Non-Dispersive Approximations for Transient Ray Fields in an Inhomogeneous Medium," to be published in the Proceedings of the NATO Advanced Research Workshop on Hybrid Formulation of Wave Propagation and Scattering (L.B. Felsen, editor) Nijhoff Publ. Co., The Hague, Netherlands.

10. L.B. Felsen and A. Kamel, "Hybrid Ray-Mode Formulation of Parallel Plane Waveguide Green's Functions,"* IEEE Trans. on Antennas and Propag. AP-29, pp. 637-649, 1981.

11. L.B. Felsen, "Hybrid Ray-Mode Fields in Inhomogeneous Waveguides and Ducts," J. Acoust. Soc. Am., 69 (2), pp. 352-361, Feb. 1981.

12. A. Kamel and L.B. Felsen, "Hybrid Ray-Mode Formulation of SH Motion in a Two-Layer Half Space," Bull. Seismol. Soc. Am., 71, 6, pp. 1763-1781, 1981.

---

* Best Paper Award for 1981, IEEE Antennas and Propagation Society.

13. E. Niver, S.H. Cho and L.B. Felsen, "Rays and Modes in an Acoustic Channel with Exponential Velocity Provile," Radio Science, Vol. 16, No. 6, pp. 963-970, Nov.-Dec., 1981.

14. E. Topuz, E. Niver, and L.B. Felsen, "Electromagnetic Fields Near a Concave Perfectly Conducting Cylindrical Surface," IEEE Trans. on Antennas and Propag., Vol. AP-30, No. 2, pp. 280-292, March, 1982.

15. F.B. Jensen and W.A. Kuperman, "Sound Propagation in a Wedge-Shaped Ocean with a Penetrable Bottom," J. Acoust. Soc. Am. 67, 1564-1566 (1980).

16. A. Kamel and L.B. Felsen, "On the Ray Equivalent of a Group of Modes," J. Acoust. Soc. Am., 71, pp. 1445-1452, 1982.

17. C.G. Migliora, L.B. Felsen and S.H. Cho, "High-Frequency Propagation in an Elevated Tropospheric Duct," IEEE Trans. on Antennas and Propag. AP-30, pp. 1107-1120, 1982.

18. A.D. Pierce, "Augmented Adiabatic Mode Theory for Upslope Propagation From a Point Source in Variable-Depth Shallow Water Overlying a Fluid Bottom," J. Acoust. Soc. Am. 74, 1837-1847 (1983).

19. P.D. Einziger* and L.B. Felsen, "Rigorous Asymptotic Analysis of Transmission Through a Curved Dielectric Slab," IEEE Trans. on Antennas and Propag., AP-31, No. 6, pp. 863-870, Nov., 1983.

20. P.D. Einziger* and L.B. Felsen, "Ray Analysis of Two-Dimensional Radomes," IEEE Trans. on Antennas and Propag., AP-31, No. 6, pp. 870-884, Nov., 1983.

21. A. Kamel and L.B. Felsen, "Hybrid Green's Function for SH Motion in a Low Velocity Layer," Wave Motion 5, pp. 83-97, 1983.

22. I.T. Lu, L.B. Felsen and A. Kamel, "Eigenreverberations, Eigenmodes and Hybrid Combinations - A New approach to Propagation in Multiwave Layered Media," to be published in Wave Motion.

23. J.M. Arnold and L.B. Felsen, "Ray Invariants, Plane Wave Spectra, and Adiabatic Modes for Tapered Dielectric Waveguides," to be published in Radio Science.

24. Felsen, L.B., Numerically Efficient Spectral Representations for Guided Ocean Acoustics. Invited paper, Workshop on Computational Ocean Acoustics, Yale University, Aug. 1-3, 1984.

---

* R.W.P. King Award for 1983, IEEE Antennas and Propagation Society.

E. HIGH POWER MICROWAVE-ATMOSPHERE INTERACTIONS

Professors N. Marcuvitz and W.T. Walter

Unit EM4-5

1. OBJECTIVE(S)

Study of high power microwave-matter interactions in the atmosphere in order to obtain a detailed space-time dependent description of the nonlinear and turbulent processes involved in the conversion of microwave energy into particle dynamic energy. Of applicational concern will be the determination of the increase of pressure and temperature that can be produced at different altitudes as a function of power level, pulse length, repetition rate, frequency, focusing configuration, etc.

This study continues the basic approaches developed earlier by us for understanding nonlinear and turbulent processes in wave-matter interactions. The primary concern here is with interactions between high power microwaves and the atmosphere, and with associated modular problems dealing with wave propagation and absorption.

2. APPROACH

The approach to the overall space-time description of high power microwave absorption in the atmosphere is based on a number of past and current theoretical and experimental wave-matter activities of faculty and graduate students working on nonlinear and turbulent wave propagation at both microwave and laser frequencies. Our approach is not to evolve a massive "code" for computation, but rather to concentrate on the analysis, in a modular sense, of those finite algorithmic models that affect the accuracy of existing procedures used in large overall codes, with particular emphasis on hydrodynamic processes that occur in the absorption of high power microsecond microwave pulses.

3. SUMMARY OF RECENT PROGRESS

1) Theoretical

Calculations of the interaction of a high power modulated microwave pulse with the atmosphere have continued. Major emphasis in the past period was placed on determining the accuracy of the quasiparticle method, and this has resulted in a novel approach to obtaining exact solutions of wave propagation problems. As described in our previous reports, the interaction of a microwave pulse with an ionized atmosphere requires a self-consistent description of both ionized particles and propagating wave pulses. By viewing localized wave pulses as systems of quasiparticles, one reduces the overall pulse-atmosphere problem to that of the interaction of different "particles"--the electrons, ions and neutrals characteristic of the ionized atmosphere, and the quasiparticles representative of the electromagnetic wave excitations. Kinetic equations are employed to describe the dynamics of these interacting "particle" systems. The derivation of "exact" kinetic equations to describe real particles in the presence of processes involving collisions, ionization, attachment, etc., has been discussed in previous reports and in the literature. The derivation of exact quasiparticle kinetic equations is also well known[3] but solutions have been obtained only to first or second order of approximations. The accuracy of these approximations is limited by the presence of rapid space-time variability in the field or medium. Singular caustic phenomena occur in such regions, and conventional wave propagation calculations via ray, normal mode, or saddle point techniques become cumbersome or fail. In the past period we have evolved an exact trajectory based method of solving quasiparticle kinetic equations; a preliminary report on this work will be presented[28] at the September 1984 URSI General Assembly in Florence, Italy.

To summarize briefly this recent development, we consider the description of the field radiated by an antenna feed into the near-to-far-zone transition region in front of the feed; this region poses both analytical and computational difficulties in conventional calculations. Although this modular problem considers free space propagation in the absence of an ionized atmosphere, it does provide an indication of the effects of the exact treatment of transition regions, wide angle propagation, etc. Four displays are presented here depicting the quasiparticle field view at $z = 1$ for the case of a Gaussian ($N = 1$) antenna feed distribution at $z = 0$; the choice $z = 1$ is in the transition region and corresponds to a scale parameter $1/h = k_0 d = 2\pi d/\lambda = 3.14$, d being the width of the feed and λ the wavelength. Figure 1 depicts the quasiparticle phase space density in the narrow angle first order approximation. Figure 2 displays the quasiparticle phase space density again in first order but valid for wide angles. Figure 3 displays the "exact" quasiparticle phase space density calculated numerically via a trajectory based solution of the exact kinetic equation. It is of interest to note the oscillatory and attenuated nature of the exact quasiparticle phase space density, a feature which is indicative of the generalized Airy function behavior characteristic of higher order effects. Figure 4 shows the field amplitude patterns in dB calculated from the above three quasiparticle densities. Although the display corresponds to a Gaussian excitation, $N = 1$, the program which utilizes FFT based algorithms is also valid for $N = 2,3,\ldots$etc. Analytical and numerical advantages of this quasiparticle method stem from the relatively slow space time variations of quasiparticle measures of wave fields as compared to those employed in conventional field descriptions. As previously noted, the method is generally applicable to linear, nonlinear, and turbulent wave propagation.

2) <u>Experimental</u>

The magnetron previously utilized in our experimental verification studies of our pulse propagation model had a peak power capability of 50 kW at 3 GHz and pulse widths of 80 or 400 nanoseconds. As previously reported, pulse shortening was observed within a pressure region of 2 to 20 torr around the Paschen minimum. We have noted that higher peak powers and longer pulses would be desirable to enlarge the pulse shortening region for comparison with the model. Since a 5795 magnetron was available in our storeroom with a 1 MW peak power capability, we have constructed a new high-power modulator with sufficient capability to drive megawatt magnetrons and with appropriate interlock protection both for the tube as well as for the protection of

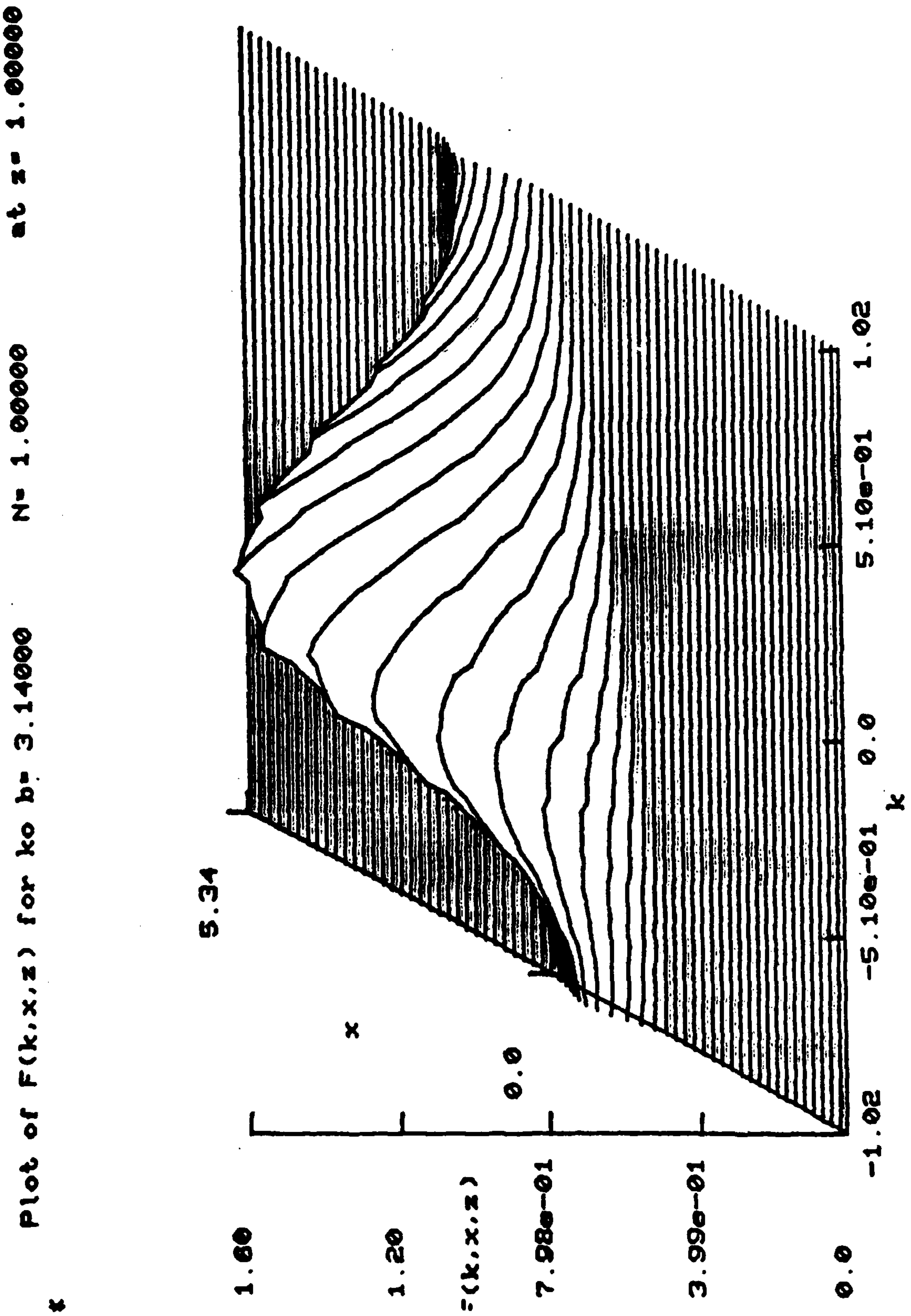


Fig. 1. Narrow angle, 1st order quasiparticle approximation.

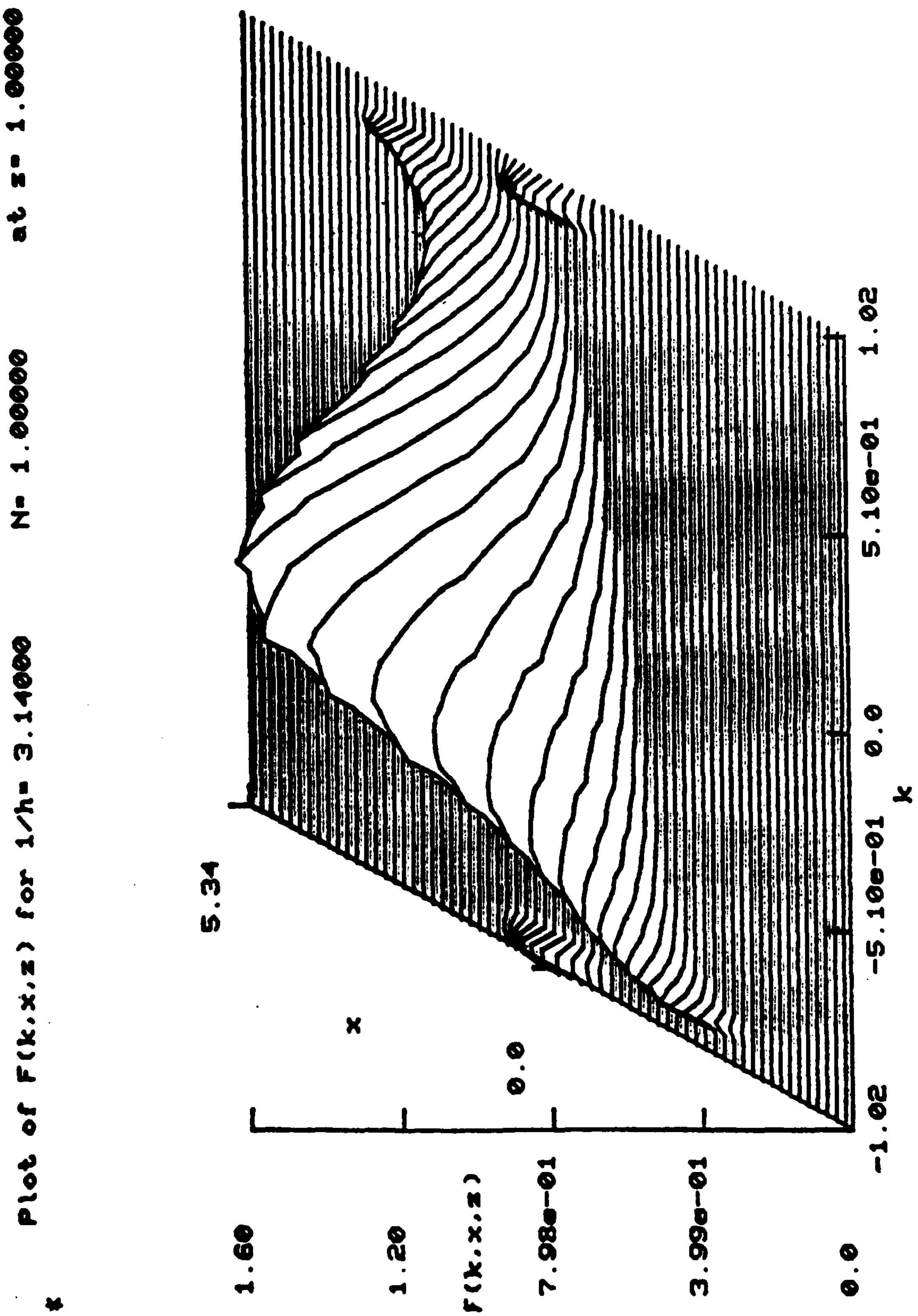


Fig. 2. Wide angle, 1storder quasiparticle approximation.

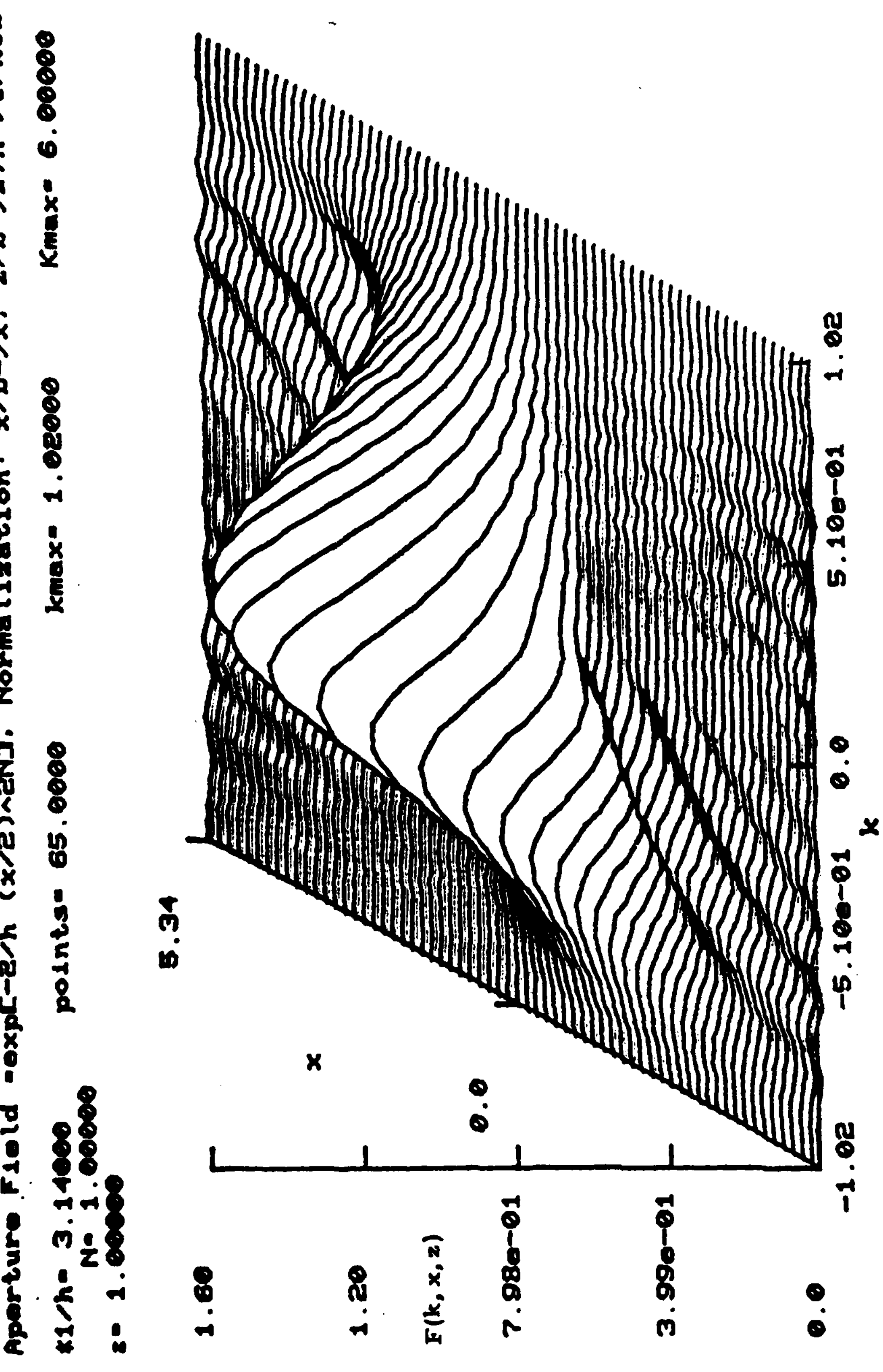


Fig. 3. Exact phase space density.

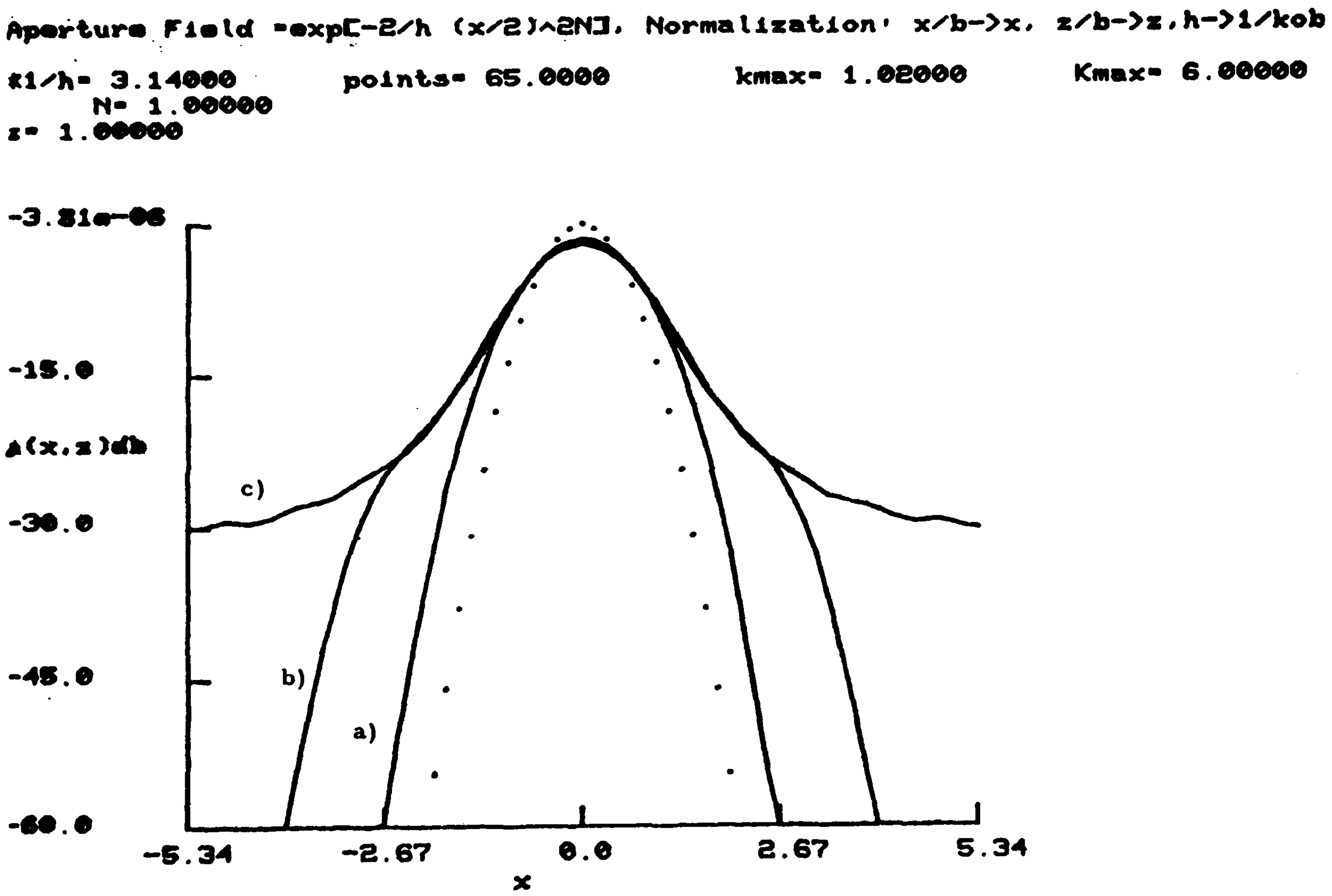


Fig. 4. a) narrow angle, b) wide angle, c) exact.
Dots represent $z=0$ distribution.

personnel. Upon completion of the modulator and testing of the magnetron, the 5795 tube was found to be gassy and unable to hold the high voltage. We wish to acknowledge the generous assistance of Raytheon Corporation. Through the assistance of Senior Engineer Richard Y. Clark we were able to obtain a 1448 magnetron from Raytheon.

When the 1448 magnetron was installed and operated on our S-band line, breakdown occurred at the input window. A pair of brass papers were constructed to provide a window location within the waveguide of larger cross-sectional area and therefore lower power/cm² loading. Installation of a mylar window between the papers did successfully overcome the problem of breakdown at the input window. Three detailed series of pulse shortening measurements were taken as a function of pressure within the S-band waveguide at 20, 24, and 29 kV, corresponding to peak powers of 225, 400, and 600 kW. At higher magnetron voltages, continuous breakdown took place. Shortening of the 3 μsec pulses began at 81 torr for the 225 kW pulse, at 102 torr for the 400 kW pulse, and at 151 torr for the 600 kW pulse. In each case, the lower pressure recovery of the full pulse width occurred at pressures below 0.2 torr, which was the lowest pressure attainable with a Welch Model 1397 mechanical pump attached to our S-band waveguide system. These measurements provide an extended data base for comparison with our computer models of microwave breakdown. Because of the departure of Dr. W. Walter from the Polytechnic, and the expenses involved in the running of the high power microwave experiment, we have terminated this phase of the program.

3) Computational Facilities

To overcome the twin problems of additional addressable space and increased speed, we utilized equipment funding from JSEP during 1983 to order a CSPI array processor on a Gould 32/6750 super mini-computer. The host Gould computer has a 32-bit word structure, 2 megabytes of fast memory and is about six times faster than the PDP 11/60. In addition, the CSPI 6410 array processor, which has 32K words of 32-bit program/data memory and 64K words of 64-bit data memory, can provide additional speed enhancements of up to 200 depending on the program. We are hoping for a speed enhancement of 10 to 20 which would then permit extension of our computer modeling results to longer pulses. The Gould 32/6750 and CSPI 6410 array processor arrived at the Polytechnic in September, 1983.

John Buck has completed a C-compiled rewrite of our interactive graphic language IGL used for our computer model algorithms; this IGL version runs 5-6 times faster on the Gould than on our PDP 11/60. Using the CSPI array processor for computer-based sector or matrix array calculations improves the speed by a factor of about 20 faster than the 11/60.

4. STATE OF THE ART AND PROGRESS DETAILS

The current and anticipated availability of very-high-power microwave sources is leading to a renewal of interest in effects caused by dumping of large amounts of energy into atmospheric regions that have been broken down and ionized. The determination of the resulting enhancement of pressure, temperature, etc., of the atmospheric constituents requires a self-consistent space-time dependent description of the electromagnetic fields and of the electron, ion, and neutral particle dynamics within the interaction region. Of particular interest are effects arising from power levels above breakdown (> 10^6 watts/cm^2 at ground level) and pulse lengths in the microsecond range and longer. The equations descriptive of the interaction are the Maxwell equations together with kinetic and/or fluid dynamic equations for the particle constituents with coupling represented by "collision" terms for the different wave-particle, particle-particle, and wave-wave interactions involved. Researchers have employed a variety of descriptive models that require a mix of analytical assumptions, extrapolated experimental data, and numerical computations. In the range of interest there is as yet no generally accepted objectionless description that represents an optimum and "cost effective" mix of analysis and algorithmic modeling.

Among the criticisms that have been leveled at current approaches to high power microwave-atmosphere interaction problems are:

(a) The space-time dependent absorption and reflection electromagnetics of an incident microwave pulse are not treated accurately; reflections as well as transverse to longitudinal mode coupling are either neglected completely or are taken into account by ad hoc procedures of questionable accuracy.

(b) The non-equilibrium nature of the particle kinetics is not adequately reflected in fluid descriptions of the various particle species.

(c) In the energy range of interest collision parameters representative of the atmospheric chemistry and descriptive of momentum and energy transfer, ionization, attachment, recombination, heat conductivity, etc. are derived with questionable accuracy by extrapolation from a low energy base. There is also a question as to the optimum number of chemical reactions necessary to achieve engineering accuracy.

Although the ultimate microwave pulses of interest must have peak powers in the vicinity of static breakdown (10^6 watts/cm^2 at ground level) and microsecond or longer pulse lengths, only two facilities within the United States are presently under construction to generate such pulses, one partially completed at the Naval Research Laboratory and one planned at General Dynamics. The tests scheduled for these facilities tend to be global in nature, i.e., they demonstrate <u>overall</u> effects, but do not follow <u>individual</u> interaction processes that elucidate the basic phenomena. Therefore, in view of the substantial time delay before gigawatt facilities become operational and the global character of the scheduled very high power tests, <u>experiments at lower peak power levels</u> will be highly desirable both to <u>test out the evolving theory</u> and to provide significant program direction.

Although an impressive amount of research, both analytical and simulated, has been carried out on space-time dependent descriptions of wave-matter interactions by workers in this country and by the Russians, there is comparatively little that has been done in the microwave frequency range.

SECTION I: ELECTROMAGNETICS

In the past our group at the Polytechnic has investigated high-power, short-pulse microwave propagation in the atmosphere with support from JSEP, NAV AIR and NRL. One of our contributions was the first detailed space-time analysis and computer display of high-power microwave pulse propagation in the presence of atmospheric breakdown. This point was noted during a meeting of the DoD special committee on microwave technology, on which one of our group (Professor Marcuvitz) was a member. The experience gained from these involvements in microwave absorption processes in the atmosphere has underscored the need for addressing the criticisms indicated above.

Apropos of item (a) above, we have considered the application of quasiparticle techniques to absorption and reflection of pulse modulated microwave carriers by an inhomogeneous dielectric medium; there are well known analytical difficulties in the description of the reflected pulse because of the double frequency coupling terms arising in the inhomogeneous region. We have investigated a number of different techniques for averaging out the complicated high frequency space-time dependent terms so as to retain the important low frequency terms that determine energy-momentum transport. In particular, we have calculated overall energy transport over a large parameter range by concentrating on far field reflection and transmission coefficients. The accuracy of these calculations has been checked by comparison with numerical computations, at least in the range wherein computer runs are economical in terms of memory and execution time requirements. We have also explored analytically and numerically energy conversion processes associated with the interaction of a high power modulated microwave pulse and an atmosphere ionized by the pulse. The basic problem is that of a pulsed beam of finite spatial width and temporal extent incident obliquely on an atmosphere at power levels such that the plasma formed is nonlinear, turbulent, and space-time variable. As noted above, analysis of this problem is dependent on the derivation of suitable equations descriptive of the electromagnetic and particle dynamics, and capable of being manipulated into a form for efficient numerical analysis. To obtain computational efficiency we have employed quasiparticle methods to describe both the microwave fields and the collective processes excited in the plasma formed by the pulse; these methods average out fast space-time oscillatory behavior and retain only the relatively slow space-time variability of engineering interest. As noted in Sec. 3, a number of modular problems have been considered to obtain accuracy checks on our computational algorithms via a novel exact procedure.

In connection with the item (b) above, we have investigated differences between results of kinetic and fluid model descriptions of an electron plasma under different types of excitation. It is known that collisionless kinetic models of particle systems can be approximated by unviscid nonlinear fluid models that under certain types of excitation lead to shock phenomena not present in the original kinetic model. This discrepancy, which appears to stem from the non-equilibrium nature of the particle kinetics, can be eliminated at least in certain ranges by the introduction of anomalous heat conductivity collision terms into the otherwise collisionless fluid model. However, the presence of sufficient particle-particle collisions of sufficiently large pulsed electric field excitation seems to obscure the above discrepancy. We have explored numerically parameter ranges in which the above discrepancy is or is not important.

SECTION I: ELECTROMAGNETICS

In the predecessor work unit, entitled "Wave-Matter Interactions," general approaches were developed for "collision" (interaction) terms required in the description of particle-particle, particle-wave and wave-wave interaction processes which arise at those high power levels at which ionization and related phenomena take place during the propagation of an electromagnetic wave through various media. A common underlying analytical theme has been to obtain a kinetic basis for models of "collisional" interactions. Although the research reported there has been applied mostly at laser frequencies, the general approach in terms of a quasiparticle treatment of wave propagation through nonlinear media[1,2,3] is also applicable at microwave frequencies.

We have applied this approach to specific electromagnetic wave-media interactions, such as ionospheric scintillation and turbulence,[4 8,26] laser generation and propagation,[9 12] interactions with metallic surface,[13 19,22] to high-power microwave propagation and breakdown,[20,21] and in the past period to a number of related applications.[23 27] Investigation of the basic interactions in different electromagnetic frequency regimes has been helpful in the development of both analytical treatments and computer models for comparison with experiment.

To validate the microwave propagation computer model, we have assembled a conventional S-band rectangular waveguide system at the Polytechnic. A section of waveguide between two mylar windows is connected to a vacuum pump so that gas composition and pressure can be varied as desired. Air breakdown and pulse shortening are observed when the pressure within this waveguide section is reduced from atmospheric to a region in the vicinity of the Paschen minimum. In past reports we have presented a large number of experimental data obtained with this facility, including comparisons with corresponding theoretical computer models.

5. REFERENCES

1. N. Marcuvitz, "Quasiparticle View of Beam Propagation in Nonlinear Media," Nonlinear Electromagnetics, ed. P. Uslenghi, Academic Press, p. 117-132 (1980).

2. N. Marcuvitz, "Nonlinear Wave Propagation," URSI International Symposium on Electromagnetic Waves, Munich, Germany (September 1980). Invited.

3. N. Marcuvitz, "Quasiparticle View of Wave Propagation," Proc. IEEE, Vol. 68, p. 1380 (November 1980). Invited.

4. S.R. Barone, "Nonlinear Theory of Type II Irregularities in the Equatorial Electrojet," Phys. of Fluids, Vol. 23, p. 491 (March 1980).

5. S.R. Barone, "Renormalization of Maxwell's Equations for Turbulent Plasma," Phys. Rev. A, Vol. 21, p. 1725 (May 1980).

6. D.M. Wu, "Ionospheric Scintillation," Ph.D. Thesis, Polytechnic Institute of New York (August 1980); also POLY-MRI Report No. 80-1410.

7. S.R. Barone, "Angular Momentum of Particles in Geomagnetic Field," Phys. Letters, Vol. 80a (December 4, 1980).

8. N. Marcuvitz and D. Wu, "Ionospheric Scintillations," URSI Boulder Meeting (January 1981).

9. M. Newstein and F. Mattar, "Adaptive Stretching and Rezoning as Effective Computational Techniques for Two-Level Paraxial Maxwell-Bloch Simulation," Computer-Phys. Commun., Vol. 20, pp. 139-163 (1980).

10. G.M. Kull, "Analysis of the Copper Vapor Laser and Development of a General Laser Plasma Model," Ph.D. Thesis, Polytechnic Institute of New York (May 1980).

11. W.T. Walter, N. Solimene and G.M. Kull, "Computer Modeling to Direct Copper Vapor Laser Development," Proc. of International Conference on LASERS '80 (C.B. Collins, editor); STS Press: McLean, Virginia, pp. 148-158 (1981).

12. W.T. Walter, "Stepwise Excitation - A Limiting Process in the Copper Vapor Laser," Proc. of International Conference on LASERS '81 (C.B. Collins, editor); S.T.S. Press, McLean, Virginia, pp. 853-865 (1982).

13. W.T. Walter, "Reflectance Changes of Metals During Laser Irradiation," Proc. of SPIE, Vol. 198, ed. J.R. Ready and C.B. Shaw, Jr., pp. 109-117 (1980).

14. N. Solimene and M. Newstein, "Laser-Metal Interaction in Vacuum," XI International Quantum Electronics Conference, Boston (June 1980).

15. T.H. Kim, "High-Energy Pulsed Laser Interaction with Metallic Surfaces," Ph.D. Thesis, Polytechnic Institute of New York (June 1980).

16. M. Newstein and N. Solimene, "Laser Metal Interactions in Vacuum," to appear in IEEE J. of Quantum Electronics, Vol. QE-17, pp. 2085-2091 (October 1981).

17. M. Newstein, "Laser Interaction with Metals," Proc. of International Conference on LASERS '80 (C.B. Collins, editor); STS Press: McLean, Virginia, pp. 71-80 (1981). Invited Paper.

18. W.T. Walter, N. Solimene, K. Park, T.H. Kim and K. Mukherjee, "Optical Properties of Metal Surfaces During Laser Irradiation," in <u>Lasers in Metallurgy</u> (K. Mukherjee and J. Mazumder, editors) TMS-AIME Conference Series, pp. 179-194 (1982). Invited Paper.

19. K. Mukherjee, T.H. Kim and W.T. Walter, "Shock Deformation and Microstructural Effects Associated with Pulse Laser-Induced Damage in Metals," in <u>Lasers in Metallurgy</u> (K. Mukherjee and J. Mazumder, editors) TMS-AIME Conference Series, pp. 137-150 (1982). Invited Paper.

20. N. Marcuvitz, N. Solimene and W.T. Walter, "High-Power Microwave Propagation," POLY-MRI Report No. 1413-80 (December 1980).

21. W.M. Bollen, W.M. Black, R.K. Parker, W.T. Walter and R.T. Tobin, "High Power Microwave Breakdown of Gases," Plasma Physics Division of American Physical Society, San Diego (November 1980).

22. W.T. Walter, N. Solimene, K. Park, T.H. Kim and K. Mukherjee, "Optical Properties of Metal Surfaces during Laser Irradiation," Proc. of International Conference on LASERS '81 (C.B. Collins, editor); STS Press, McLean, Virginia, pp. 510-524 (1982).

23. G. Kopcsay, "Propagation of Pulses in Nonlinear and Turbulent Plasma," Ph.D. thesis in preparation.

24. S.Y. Lee and N. Marcuvitz, "Beam Reflection from Lossy Dielectric Layers," Journ. of Optical Society, Vol. 73, No. 12 (December 1983).

25. S.Y. Lee and N. Marcuvitz, "Quasiparticle Description of Pulse Propagation in a Lossy Dispersive Medium," Trans. on Ant. and Prop. (April 1984).

26. D. Wu and N. Marcuvitz, "Ionospheric Scintillations," Radio Science (October 1984).

27. N. Marcuvitz, "Quasiparticle Method for Propagation in Inhomogeneous Media," Proc. of the 1983 URSI International Symposium on Electromagnetic Theory, pp. 33-36, Santiago de Compostela, Spain (August 23-26, 1983).

28. N. Marcuvitz, "Exact Kinetic and Fluid Representations of Wavepacket Propagation," presented at XXIst URSI General Assembly, Sept. 1984, Florence, Italy.

SECTION II

SOLID STATE ELECTRONICS

SECTION II: SOLID STATE ELECTRONICS

A. ENHANCEMENT OF INELASTIC OPTICAL PROCESSES ON SMALL AND ULTRA-SMALL SOLID STRUCTURES

Professors S. Arnold, H. J. Juretschke, A.A. Oliner and P.S. Riseborough

Unit SS4-1

1. OBJECTIVE(S)

The initially-astonishing phenomenon of surface-enhanced Raman scattering (SERS) was discovered about three years ago, when it was found that the Raman scattering from selected active molecules was increased by seven or so orders of magnitude when the molecules were placed on a rough silver surface. A large number of experimental, and some theoretical, investigations followed; although there is still disagreement with respect to the physics of the effect, most researchers believe it is due primarily to surface plasmon resonances of microscopic metal particles in the optical range, even though full understanding has obviously not been achieved. Very recently, it is being recognized that similar spectacular enhancements can be achieved for a large variety of other inelastic optical phenomena, such as fluorescence, photoemission, second harmonic generation, photocatalysis, photophoretic forces, and so on. In addition, specialized geometries, such as periodic arrays, are being proposed to deliberately sharpen the resonances and improve the efficiency of the process. We believe that we are on the threshold of an exciting new field, with enormous potential for both the production or enhancement of new inelastic optical effects and the ability to exploit in various ways the resonances which produce these enhancements.

Our broad objective is to explore, understand, and optimize these enhancements of inelastic optical processes on small and ultrasmall structures through the use of well-defined geometrical configurations where experiment and theory allow the identification of the contributions to the local field buildups, among other mechanisms.

2. APPROACH

We are proposing a combined experimental and theoretical investigation of the fields and enhancements in a number of optical processes at the surfaces of small spherical particles, of plane surfaces with controlled superimposed roughness, and of non-spheroidal structures, and of their arrays. The experimental probes will include fluorescence, second harmonic generation and photoemission; we do not plan to do Raman scattering.

For the measurements on small spherical particles, we plan to use a novel facility that can levitate a single small spherical particle, and which has a sensitivity that can detect the departure of a single electron and a stability whereby the particle's charge value can remain constant within an electron for as long as a day. The experiments involving controlled roughness will take advantage of the thin film facilities and expertise that apply to Work Unit SS3-2. The theoretical modeling will concentrate on useful and clear formulations of electromagnetic field resonances in such geometries and their extensions to array and other environments, and on quantum-mechanical contributions to such local fields.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

The program on microstructure resonances has concentrated in this period on preliminary work in anticipation of extending our studies to very small-size objects and of exploring other consequences of the surface enhanced fields at any one of the resonances.

We have developed a novel fully automated levitation system, using a quadrupolar geometry and electrodynamic levitation with feedback provided by a position sensitive detector, which should enable us to study single particles in the 100 Å range. We have also carried out theoretical work on the resonances to be expected in such very small objects where the dielectric response may differ from its bulk behavior.

Finally, tapered cylindrical fibers have been successfully explored as a means of providing a continuously variable size parameter in order to separate resonating and nonresonating regions at a fixed frequency. The observed resonances are very sharp, and are suitable for use as source fields for studying the behavior of transitions in molecular systems.

In connection with the electromagnetic interactions at perfect and near perfect surfaces, the second harmonic of 1 micron radiation has now been observed on pure and flat single crystal Ag surfaces, under nonresonant conditions. The small absolute value of the coupling under these conditions remains to be determined. However, there are also preliminary indications of structural effects relating to the bonding of silver atoms in the first few layers of the surface.

4. STATE OF THE ART AND PROGRESS DETAILS

A. General Background

The recent finding that the intensity of Raman scattering in certain molecules is enhanced by many orders of magnitude (SERS) when the molecule sits close to a solid surface having particular structural characteristics has let loose a flood of investigations in order to probe a great variety of different experimental aspects of this phenomenon, and to build theoretical models for explaining these spectacular observations. Without going into detail in either experiment or theory, two general points are clearly emerging from this work. First of all, this effect is only one of many possible optical interactions taking place at surfaces that will show considerable enhancement. Secondly, by effectively extending the range of the strong coupling of electromagnetic waves with structures to a domain where the physical size of the structure becomes small compared to the wavelength, such interactions are bound to lead to novel practical applications involving various aspects of electromagnetic radiation. With the perspective of these two points in mind, we propose here to carry out a new study in this field. It builds on our considerable backgrounds in both electromagnetic wave interactions and in surface properties gained in part through past and current JSEP programs, as well as on a novel and highly sensitive measurement technique for probing optical effects at well-defined surfaces.

The enhancements (of 10^6 or so) observed in SERS are widely documented, though often only in sketchy form and with only rather qualitative background information, but even review articles[1,2] can hardly keep pace with the new findings. The probing of enhancements in other optical interactions is only beginning. To site a few significant examples, there is the 100-fold enhancement of photoemission yield from less than 50 Å diameter silver particles observed in a sol,[3] the second harmonic generation, of enhancement 10^4, from metal surfaces chemically roughened to a depth of 500 Å,[4] and the observation of strong luminescence of dye molecules on metal islands, in contrast to the quenching of this radiation on flat metal surfaces.[5]

Many of these results are not fully understood, but it is usually agreed that one or more of the following factors must play a role. Firstly, the proper local electromagnetic fields consistent with the nearby boundary conditions must be used for both the exciting and scattered radiation.[6,7] This is particularly important when either or both fields have frequencies that can excite electromagnetic modes in the solid related to geometrical resonances. Secondly, both fields may be subject to resonances in the dielectric response of the surface, because of coupling to other excitations in the solid, such as extended surface plasmons, made possible by the surface structure.[8] Thirdly, the microscopic response of the surface, i.e., the form of the quantum-mechanically self-consistent electromagnetic fields within and close to the surface resulting from spatially-varying and non-local dielectric properties of the surface region, can be a sensitive function of the surface structure.[9] Finally, there are arguments that some enhancement is connected with special active sites of atomic dimensions.[10]

Understanding the relative role of these various contributions is essential if one wants to assess the general potential for enhancing inelastic optical phenomena by the use of structural surface features. We propose a study that, on the one hand, will build on canonical surface configurations to separate and identify various contributions to structural enhancement and, on the other, will develop and apply general criteria for utilizing the method for investigating otherwise very weak higher-order optical interaction effects. The work will be both experimental and theoretical.

The two surface configurations that form the starting point of the experimental program rely on unique expertise available in our laboratory. The surfaces are (a) single spheres of dimensions comparable to and smaller in size than the wavelength of light, and (b) nearly perfect, flat single crystal metallic surfaces. Professor Arnold, who is currently an Alfred P. Sloan fellow, has perfected novel techniques for the stable suspension of single very small particles in a configuration that readily lends itself to investigating optical effects on their surfaces.[11] The extraordinary sensitivity of his instrumentation expressed in terms of current, is of the order of one electron per day, and with it he has already demonstrated the detection of striking effects on local optical fields with changes in particle size as small as a few angstroms.[12] These spheres provide a well-defined and theoretically-tractable geometry which allows unambiguous and quantitative evaluation of both classical and quantum mechanical influences on the local optical field. This is in contrast to the usual experiments on rough surfaces where an undefined fraction of the surface is active, and under only grossly-defined geometrical conditions.

Professor Juretschke has specialized in the preparation of metallic surfaces that are flat and perfect to the extent of being practically inert to interaction with an environment. Such surfaces will serve as a starting point for studying the effects of the controlled and deliberate introduction of structural defects, ranging from point sites to surface dislocations, to hillocks and periodic modulations. Again, the initial geometry is well defined, and the roughening can be <u>controlled</u>.

The theoretical effort is expected to be in strong support of the experimental program. It must concentrate on both the macroscopic and microscopic properties of any such structures, and on the local fields that are in consonance with these properties. It has an equally natural strength in this laboratory. Professor Oliner brings substantial insight and elegant mathematical techniques to the understanding of local electromagnetic fields in complex geometries. This will complement Professor Riseborough's wide background in many-body theory and transport phenomena -- he recently came to the Physics department from the University of California at Santa Barbara -- for studying their microscopic dielectric response, taking quantum-mechanical effects into account.

As already mentioned, this program has two central themes.

(a) The unambiguous determination of the contribution of local field buildups to the observed enhancements, and the theoretical modeling of some other contributions,

(b) The development of general criteria for optimal enhancements of weak higher-order optical phenomena.

Within this program, the work will have four main thrusts:

1) Resonances and resonant optical effects at the surface of single nonmetallic particles (Arnold, Oliner, Riseborough)

2) Optical interactions at surfaces with controlled roughness (Juretschke, Arnold)

3) Electromagnetic theory of plasmon resonances on small non-spheroidal particles, including the effect of array and other environments (Oliner, Juretschke, Arnold)

4) Quantum-mechanical aspects of local electromagnetic fields close to metal surfaces and small structures (Riseborough, Juretschke)

SECTION II: SOLID STATE ELECTRONICS

This division is to some extent purely formal, and does not at all imply separate efforts by the different members of the work unit. As specified above, it is, in fact, expected that the main work within any subdivision will involve a very close collaboration by at least two of the four investigators, with the pairing shifting for the different subunits. Other interactions will, of course, develop naturally as the work proceeds.

The four main areas of research will be discussed in a general way below, and the detailed program proposed in each area is presented in a subsequent section.

B. Resonances and Resonant Optical Effects at the Surface of Single Nonmetallic Particles

(1) Introduction

A single sphere is undoubtedly the simplest small structure for which the interaction with an electromagnetic plane wave can be fully specified. It is therefore also an ideal geometry in which to search for contributions to optical interactions on its surface that go beyond those arising from a purely electromagnetic field buildup. Before reaching that stage, however, a number of both experimental and theoretical questions must be clarified. For example, from the experimental side it has to be established that the effects of molecules covering only a fraction of the surface of a very small particle are, in fact, detectable with any kind of sensitivity. Secondly, one also requires to have good first-hand information about the optical properties of the bare particle itself, in order to be able to interpret the results in a meaningful way. On the theoretical side, it is equally important to have quasi-intuitive guidelines for understanding the change in the response of the sphere as the various parameters of the problem of the interaction with electromagnetic waves are changed, above and beyond obtaining the complete but usually not very transparent, and entirely computer-dependent, formula solutions. We will outline here some answers to these questions, show that an approach along these lines will, in fact, be feasible, and thereby lay the groundwork for more detailed investigations. It will become clear, though, that these methods alone open up new ways of determining the properties of small particles that remain to be exploited in their own right. We will concentrate on discussing nonmetallic particles, although many of the points also apply to semiconductors or metals.

While the straightforward detection of molecules on the surfaces of small dielectric particles is limited by their extremely small number, local resonances can improve the response appreciably. For example, in the case of dielectrics, enhancements can exist due to electromagnetic surface wave resonances. Recently, Chang et al.[13] have detected spikes in the external excitation wavelength dependence of emitted fluorescence from dye-coated optical fibers that correlate well with the surface wave resonant modes of these structures. Similar enhancements due to resonances are predicted for the emissivity spectrum of carbon-coated alumina spheres. According to Pluchino,[14] a 10Å thick outer layer of carbon should increase the emissivity of a 6 micron alumina sphere by three orders of magnitude. Pluchino feels that surface waves are responsible for this enhancement; however, there are no substantiating experiments for his prediction or for the proposed mechanism.

A direct method for verifying this would not be easy. Unfortunately it is difficult to quantify fluorescence experiments, because of the sparse area in which fluorescence photons fall, and because the precise way in which an excited molecular adsorbate interacts with a curved dielectric surface is unknown. Another and a better way to arrive at a quantitative understanding of such effects is through a thermal spectroscopy, since the partition of thermal energy between a particle and the outside environment is easily analyzed.

In addition to providing the proper tool for such thermal spectroscopy, the general technique making it possible is much more versatile, and also lends itself to probing a number of other small particle properties that are of interest to this work.

In light of the above, our objectives are two-fold:

(a) To establish new experimental methodologies for quantifying the interaction of electromagnetic waves with small particles, and with molecules on the surfaces of these particles.

(b) To use our traditional strength in electromagnetic theory at the Polytechnic in order to evolve a physical understanding of the numbers arrived at in (a).

(2) Experimental Methodology

In order to understand the manner in which dielectric spheres adsorb radiation we have established a unique probe. This probe obtains its information from the radiometric force on an illuminated particle in a gaseous atmosphere. This radiometric force, known as the photophoretic force, was discovered in 1917 by Ehrenhaft.[15] At a pressure for which the molecular mean-free-path is smaller than the particle size, the photophoretic force F_{ph} is given by

$$\vec{F}_{ph} = - CRJ\vec{I} \tag{1}$$

where C is the quantity which depends on well-known thermal and hydrodynamic properties of the particle and the gas, R is the radius of the particle, $\vec{I}$ is the incident light intensity, and J is a measure of the anisotropy of the internal heat sources. In particular[16]

$$J = Q_{abs} \; \frac{3}{8} \; \frac{\int_0^1 dx' x'^3 \int_0^\pi d\theta \sin 2\theta \, B(x',\theta)}{\int_0^1 dx' x'^2 \int_0^\pi d\theta \sin\theta \, B(x',\theta)} \tag{2}$$

where Q_{abs} is the efficiency for absorption, and $B(x',\theta)$ is the square of the electric field density at coordinate (x',θ) within the sphere (x' is a reduced radius, i.e., $x' = r/R$). The coordinate system for using

Eq. (1) is shown in Figure 1. As one can see, the sign of J is controlled by the expression in the brackets. If the internal heat sources, that are proportional to B, are primarily on the side away from the excitation, J will be positive. In the opposite case, J will be negative.

For a small quantity of surface adsorbate, the angular distribution of sources $B(1,\theta)$ is expected to be the same as in the case without adsorbates so that a measurement of J provides a measurement of Q_{abs}. Inasmuch as emissivity and adsorption efficiency are equal through Kirchhoff's law, a measurement of J through the photophoretic force also provides a measurement of the emissivity E.

A second feature of this method is that one well-defined particle may be investigated at a time. This is accomplished by using a modified Millikan chamber as a precise force balance for the (slightly charged) particle.[11] Figure 2 shows the basic scheme of the balance.

The heart of the set-up is a modified Millikan capacitor (M). A nonuniform field is generated within this capacitor in order to create a potential well at the center of the plates. This in turn creates an unstable vertical potential, so that the particle (which is forced toward the center line) will attempt to move rapidly toward the top plate. The upward motion is overcome by readjusting the potential between the plates in accordance with the particle's position. The particle position is sensed by a position-sensitive detector (D) which responds to scattered laser radiation (L1 He-Ne, 2mW). An overall servo-loop contains the levitated particle within tight constraints such that the particle changes its position by less than a diameter. Since the position is virtually constant, any photophoretic force F_p that is induced by radiation transmitted through the bottom plate (E) will automatically be balanced by a change in electric field between the plates: $F_p = -q\Delta E$. Here q is the particle charge. Since the weight alone was originally balanced by the initial field, $m\vec{g} = -q\vec{E}_0$, we obtain the relationship

$$\frac{F_p}{mg} = \frac{\Delta E}{E_0} \tag{3}$$

The photophoretic force is therefore measured absolutely from the particle weight and the fractional change in plate voltage. Along with the intensity $\vec{I}$ and the appropriate thermal and fluid properties, J is then easily found using Eq. (1).

Some results discussed below show how the method can be applied. It should be noted that, while the charge q disappears in Eq. (3), it is obviously essential that q remain constant during the experiment. The conditions in the chamber can be arranged such that the particle charge stays stable within a single electron for as long as a day.

(3) Basic Photophoresis

A typical output of the computerized experimental analysis, directly giving the value of J, is shown in Fig. 3 for a slowly evaporating glycerol particle. The data were taken at an excitation wavelength of 10.63μ, with an intensity of 380 mW/cm². By keeping the wavelength

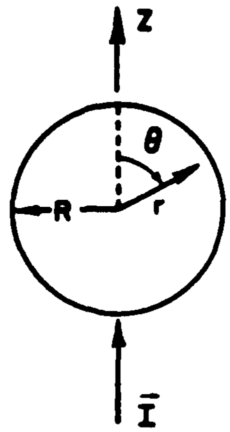


Fig. 1 Coordinate system for evaluating the integral J of Eq. (2) of the absorption of energy by a spherical particle of radius R illuminated from below with an intensity I.

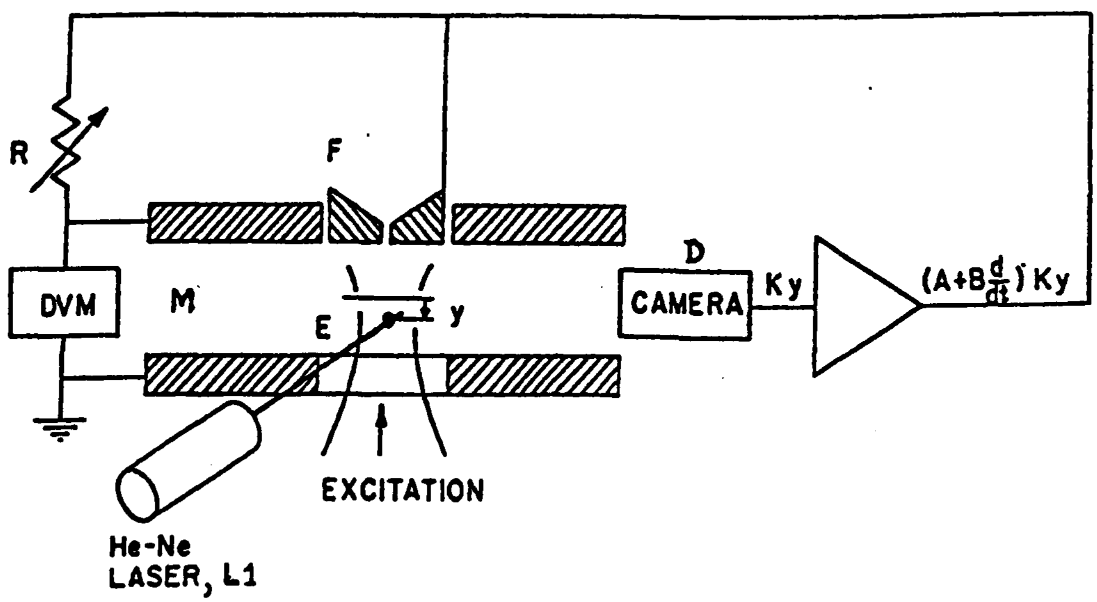


Fig. 2 Schematic of the automated Millikan chamber levitator for the stable suspension and study of single particles.

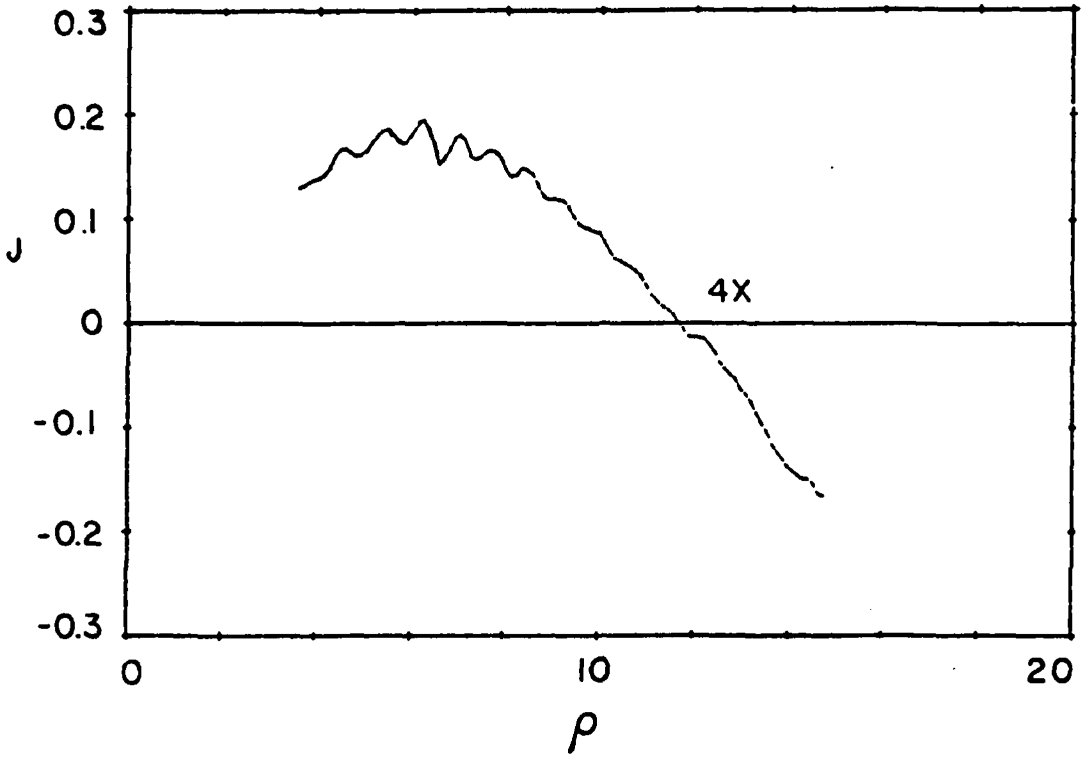


Fig. 3 The photophoretic energy absorption integral J of a slowly evaporating glycerol particle, radius R, as a function of the size parameter $\rho = 2\pi R/\lambda$.

λ constant and changing the particle size R as a function of time, we run through a continuous range of size parameter $\rho(= 2\pi R/\lambda)$, from $\rho = 15$ to $\rho = 3.5$. From $\rho = 15$ to 11.7 the force is "positive" (J is negative) and below $\rho = 11.7$ the force reverses and shows periodic structure corresponding to electromagnetic resonances. We will discuss in Section 9 why it is that the size parameter at which the force J changes sign gives the imaginary part of the refractive index, while the period of the ripples gives the corresponding real part. Hence the experiment yields directly the dielectric properties of a single small particle, a quantity that is normally very difficult to determine.

The data in Fig. 3 show that our experimental method can easily measure the quantity J. Since our primary interest is in determining the interaction of electromagnetic energy at a surface with adsorbed molecules, we must, in addition, verify that the photophoretic method can also detect a single monolayer or less.

For the purpose of obtaining such data a simple test was devised. A portion of a hydrosol of 7 micron polystyrene latex particles was deposited on a slide and dried. From the weight of this deposit the number of particles was determined. These particles were then placed in a solution of the dye Rhodamine 6G in water. The concentration of this solution was such that if the dye were fully depleted in the adsorption process it would lead to no more than ~ 2 monolayers per particle. To assure lack of clusters the solution including particles was ultrasonicated for 5 minutes. After this the entire sol was centrifuged to separate the liquid solution. This supernatant was poured off into a cuvette and tested by absorption for loss of Rhodamine 6G. It was determined that less than 5% of the dye had been adsorbed. On average, each particle had taken up less than 10% of a surface monolayer.

A single Rhodamine 6G coated polystyrene latex particle was injected into our chamber and illuminated with Ar^+ laser radiation at 5145Å and 475mW/cm^2. At a pressure of 32 torr a photophoretic force-to-weight ratio $F_p/mg = -0.015$ was formed. The negative sign indicates that the force was opposite to the direction of the radiation ("negative" photophoresis). The force measured on several particles varied by as much as 30%, presumably as a result of slightly different particle sizes and surface conditions.

If Eq. (1) is used with the constant C as prescribed in Ref. 16, and with the known quantum efficiency of Rhodamine 6G, J is found to be 6×10^{-3}. This is a very large number when one considers that the approximate geometrical absorption efficiency Q_{abs} associated with 0.1 monolayer is 3×10^{-3}. In other words, the results suggest that the asymmetric factor in brackets in Eq. (2) is approximately 2. Such a result is anomalous since the asymmetry factor in Eq. (2) cannot be larger than 1/2, and is very often a considerably smaller fraction.

One way out of this dilemma is to admit that geometrical thinking in arriving at Q_{abs} is incorrect. The geometrical picture asks, "How much light is absorbed in a single pass through the particle?" This point of view ignores surface waves which, when in resonance, can stay in the proximity of a monolayer for a much greater length of time and therefore enable Q_{abs} to be enhanced. What is required for analyzing the present experiments is a complete Mie electromagnetic scattering calculation of the associated internal fields. Experimentally, surface waves should reveal themselves through resonant structure in the photophoretic excitation spectrum, either as a function of particle radius R, or of wavelength λ. In this preliminary experiment we did not determine whether or not resonance conditions were met.

On the basis of these experimental results we have developed a program to:

(i) Obtain systematic photophoretic excitation spectra on dye-coated dielectric microspheres in order to determine the mechanisms for the large value of Q_{abs} such as seen in our preliminary results.

(ii) Analyze these experimental results via a full description for the internal fields, in order to ascertain the contributions of the local electromagnetic field buildups. Such a process (Mie theory), if carried out purely formally, often leads to very little physical insight. Therefore we construct a description which will identify the most important modal contributions to the surface fields and thereby hopefully obtain increased insight into the interaction, and especially into the conditions for optimizing it.

(4) Simple Formulation of Electromagnetic Resonances for Spheres

Since the single particles under consideration here are of spherical shape, and since we are presuming that the phenomena involved here are predominantly electromagnetic in nature, these phenomena should be characterizable in terms of the classical Mie theory[17] for the scattering of an incident plane wave by a dielectric sphere of arbitrary size. However, the Mie theory phrasing of these problems appears enormously involved, since the plane wave is geometrically not compatible with a spherical coordinate system and must therefore be expanded into an infinite set of spherical waves, and numerical results are achieved only with the help of a complicated computer program. Various publications indeed proceed in this way in their theoretical treatments (e.g., refs. 14 and 18). What is lacking is a simple and transparent theoretical formulation that yields greater physical insight by excluding the unnecessary portions of the complete Mie formalism and retaining only the essential part or parts.

We shall seek to introduce and employ such simple and transparent formulations where possible by the use of spherical transmission line theory, a rigorous technique in electromagnetics introduced in the 1950's at the Polytechnic and utilized in various contexts.[19] This approach should be particularly useful when we are interested in just the conditions for a specific resonance (and the field distribution corresponding to that resonance). Indeed, much of what we seek falls into that category. When we need the relative amplitudes of many different modes excited by a particular source, the formulation will necessarily become complicated no matter what phrasing is used.

Spherical transmission line theory is intended to parallel the common and widely-used uniform transmission line theory in its use of impedances, etc. An important similarity exists in that the modes can be divided into E modes and H modes in both representations. Important differences appear in the spherical case, however; for example, the characteristic impedance is a function of radius and is different looking towards or away from the origin. The directness of the spherical transmission line formulation may be illustrated by presenting a simple case here. Suppose we wish to write down the condition for resonance of a particular E mode; we then follow the standard transverse resonance requirement that states that

$$\overleftarrow{Y}_n(r) + \overrightarrow{Y}_n(r) = 0 \qquad (4)$$

that is, the sum of the input admittances at radius r looking towards the origin and away from it must sum to zero. The form of these admittances depends on the radial character of the spherical particle; if the sphere is simply a bare homogeneous dielectric sphere then we choose r = R, if R is the radius of the sphere, and the admittances become the characteristic admittances which, for an E mode, are

$$\frac{\overleftarrow{Y}_n(R)}{\eta_1} = -i\,\frac{\hat{J}_n(k_1R)}{\hat{J}_n'(k_1R)} \qquad (5)$$

looking towards the origin, and

$$\frac{\vec{Y}_n(R)}{\eta} = i \, \frac{\hat{H}_n^{(2)}(kR)}{\hat{H}_n^{\prime(2)}(kR)} \tag{6}$$

looking outward, away from the origin. The functions involved are the spherical Bessel functions, which are related to the ordinary Bessel functions of order (n + 1/2), and consist of combinations of sines and cosines. The prime signifies the derivative with respect to the argument,

$$\eta = \sqrt{\frac{\varepsilon_0}{\mu_0}} \quad , \quad \eta_1 = \eta \sqrt{\varepsilon_r} \quad , \quad k_1 = k \sqrt{\varepsilon_r} \tag{7}$$

where ε_r is the relative dielectric constant of the material comprising the sphere, and can be complex ($\varepsilon_r = \varepsilon' + i\varepsilon''$).

Substituting (5), (6) and (7) into (4), we obtain the rigorous resonance relation for any specific E mode as

$$\frac{\hat{H}_n^{(2)}(kR)}{\hat{H}_n^{\prime(2)}(kR)} = \sqrt{\varepsilon_r} \, \frac{\hat{J}_n(kR\sqrt{\varepsilon_r})}{\hat{J}_n'(kR\sqrt{\varepsilon_r})} \tag{8}$$

Condition (8) for the resonances of this class of spherical modes has been derived in a strikingly simple way, without all the clutter that customarily accompanies it in the usual treatments. We have no intention here of denigrating the usual treatments, which contain far more information and which aim to treat various other aspects, including the scattered far field, for which the complete treatment is required. We point out merely that when certain partial information is required, it is unnecessary to employ the complete analysis first and then extract the portion needed; one can derive the needed partial information simply and directly in many instances.

In our cases of interest, the spherical particle is metallic in the optical range and is very small, and we seek the condition for resonance of the dipole mode. That case corresponds to the solution given in the literature[18] in an explanation of the large enhancements found in SERS. The mode in that case corresponds to n = 1, so that the spherical Bessel function ratios in (8) become

$$\frac{\hat{J}_1(s)}{\hat{J}_1'(s)} = \frac{\frac{\sin(s)}{s} - \cos(s)}{\sin(s)\,[1 - \frac{1}{s^2}] + \frac{\cos(s)}{s}} \tag{9}$$

where $x = kR\sqrt{\varepsilon_r}$, and

$$\frac{\hat{H}_1^{(2)}(v)}{\hat{H}_1^{\prime(1)}(v)} = \frac{-\,i\,\frac{u}{v} - u}{\frac{u}{v} + i\,\frac{u}{v^2} + i\,\frac{1}{u}} \tag{10}$$

where $v = kR$ and $u = \exp(ikR)$.

In the limit of very small R, (9) and (10) reduce to

$$\frac{1}{2}\, kR \sqrt{\varepsilon_r} \qquad \text{and} \qquad -\, kR$$

respectively, so that (8) yields

$$\varepsilon_r = -2 \tag{11}$$

in agreement with the literature.

The advantage here is not so much that one rather readily arrives at the well-known condition (11), which shows that this resonance on a metal particle can occur only when ε_r is negative, and also tells the wavelength of resonance according to the particular metal involved. The use of (9) and (10) also permits us to find out quite directly how condition (11) changes as the particle size increases, and as the condition then becomes dependent on particle size.

The example given above is meant to be illustrative of how we hope to use spherical transmission line theory to yield needed partial information in a simple and direct way, so that numerical results may be obtained cheaply and physical insight may be more readily furnished.

(5) <u>Previous Progress</u>

The experimental program to determine the distribution of the local electromagnetic field around and within a microstructure has been developing in a number of directions.

The photophoretic determination of the anisotropy of the absorption of light within a spherical particle (the quantity J of Eq.(2), and its interpretation along the lines of Eq.(14)) has now been published.[37] In addition, these results recently received striking theoretical confirmation by A. Pluchino, who was able to locate both the oscillations and the crossover point in Fig. 3 on an absolute basis, using the complex dielectric constant of glycerol independently arrived at in Ref. 37. This confirmation assures the possibility of the complete determination of the complex dielectric constant of unknown microparticles by photophoretic measurements. From our point of view, this agreement shows that photophoresis also gives the information we are seeking concerning the field distribution within the particle.

As an interesting aside, the absolute magnitude of J as calculated is about 50% larger than the experimental value. This indicates that the actual thermal slip coefficient at the particle's surface is below the lower limit (0.75) set by classical kinetic theory, so that our data also contain information whose interpretation requires a more realistic understanding of the scattering of molecules at surfaces.

The extension of this method to probe local fields at or just above the particle's surface by using adsorbed molecules interacting with the electromagnetic field is in progress. The simplest configuration requires a known uniform coating of single molecules on the surface, and methods are being developed to deposit dye molecules without clustering and in controlled concentrations. The same theory that has confirmed the photophoretic interpretation of Fig. 3 also indicates that at resonance the internal field tends to be pushed toward the surface of the sphere. We therefore anticipate a huge enhancement in the photophoretic force for surface adsorbed molecules. Indeed, preliminary experiments indicate an enhancement of about two orders of magnitude. Such an enhancement should be accompanied by steep resonances in the force spectrum. We are currently searching for these resonances. The experiment involves the excitation of a surface coated micron-sized particle in the new quadrupole force balance. The electromagnetic theory of this enhancement models the adsorbed molecules by a uniform dielectric coating, clearly only a crude approximation to the reality of individual surface adsorbed molecules. We will let the experimental results dictate the need for a more refined model.

The understanding of our results on photophoresis, as shown in Fig. 3, has been accelerated by new theoretical developments. As already mentioned, calculations published by Pluchino,[47] based on Yalamov's theory of photophoresis,[16] reproduce the resonant features in Fig. 3, and yield a reasonable cross-over of J as a function of size. But the actual force is less than that calculated, and this discrepancy increases to over a factor of two with decreasing particle size. Although Pluchino originally attributed this discrepancy to the use of Gaussian beams in our experiment, since the theory is worked out for plane wave excitation, this supposition proved to be incorrect. We have now found, in direct collaboration with Pluchino (at Aerospace Corp.) that the correct origin of the disparity lies in the breakdown of continuum hydrodynamic theory when applied to the combination of gas pressure and particle size in our experiments. Although the mean free path in the surrounding gas is never larger than one quarter of the particle radius, the measured force is less than one half of that calculated from continuum theory. To include modifications of continuum theory that remove this discrepancy, we have evolved a <u>comprehensive new theory of photophoresis</u>.[48] The overall result of this work leads to multiplying the right side of Eq. (1) by a term $g(K_n)$ dependent on the Knudsen number K_n (ratio of gas mean free path to particle radius). The overall force then becomes

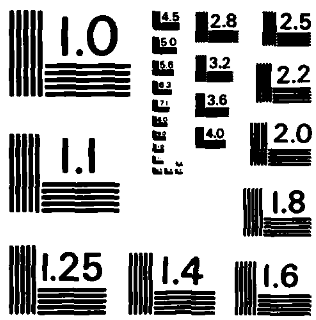
$$F_\rho = -\,c g(K_n) R J \vec{I} \tag{12}$$

with

$$g(K_n) = \frac{1}{[1+3c_m K_n]\;[1+2c_t K_n]}$$

where c_m is related to momentum accommodation at the gas-particle interface (a number experimentally between 1.00 and 1.35), and c_t is the temperature jump coefficient (a number between 1.875 and 2.48). Fig. 4 shows the J derived from the photophoretic force using Eq. (12), and compares it with the theoretical value of J. The current agreement between theory and experiment is within 10%.

AD-A147 891 BASIC RESEARCH IN ELECTRONIC (JSEP): JOINT SERVICES 2/3
ELECTRONICS PROGRAM..(U) POLYTECHNIC INST OF NEW YORK
BROOKLYN MICROWAVE RESEARCH INST.. A A OLINER
UNCLASSIFIED 30 SEP 84 POLY-MRI-1432-83 AFOSR-TR-84-0978 F/G 9/5 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

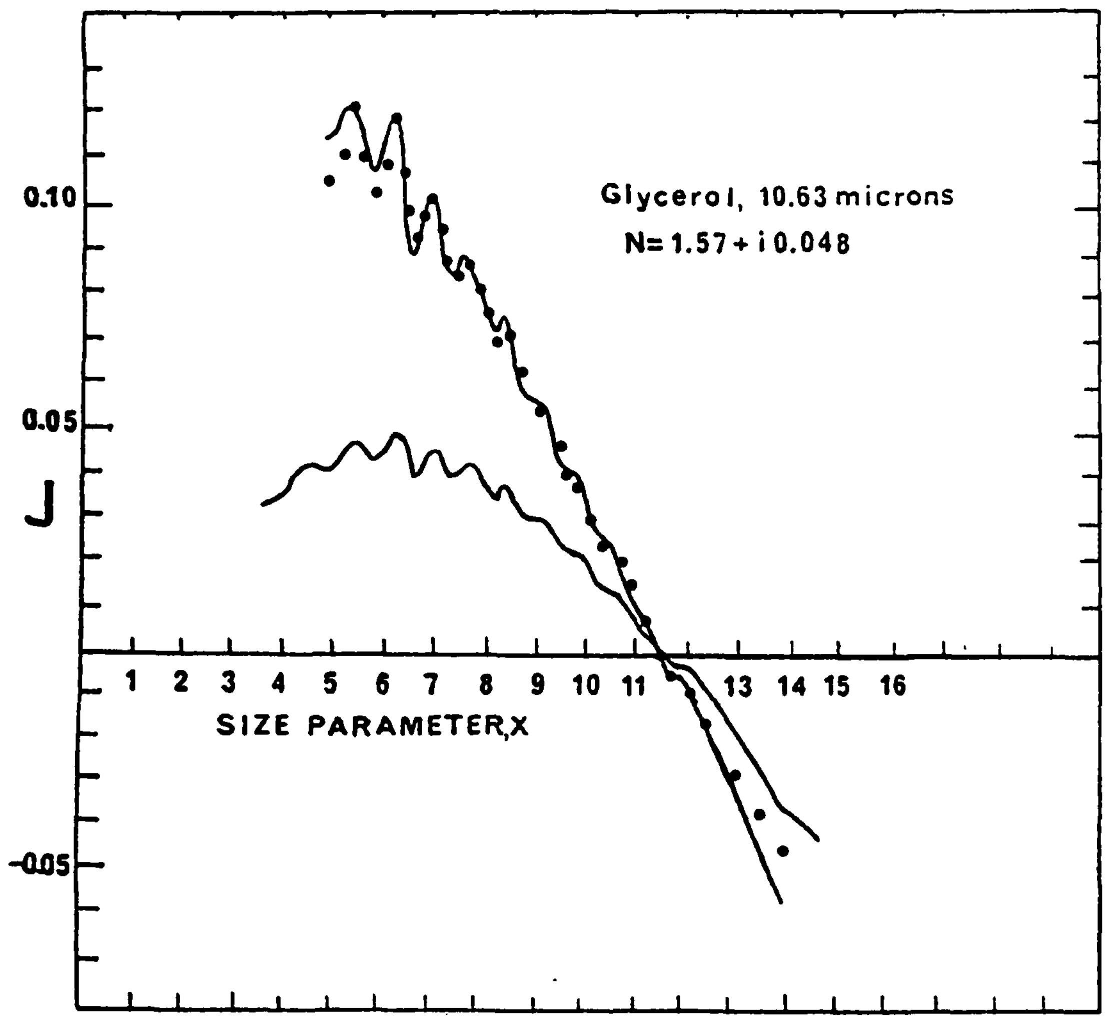


Fig. 4 Evaluation of J of Figure 3 when corrected for the breakdown of continuum hydrodynamic theory. Points show the first principles calculation of J from Mie theory.

It is interesting to note that the best fit to the data requires a thermal slip coefficient of 1.25, which is in marked disagreement with the value 3/4 predicted by Maxwell on the basis of classical kinetic theory. Our data clearly contain information whose interpretation requires a more realistic understanding of the scattering of molecules at surfaces. An improved theoretical approach to describe such scattering is currently in progress.

It is clear from our data on photophoresis that well-developed resonant modes exist within any small microsphere. Many of these modes are of sufficiently high order that calculations of the anisotropic internal field distribution require a complicated polynomial expansion, and physical insight is difficult. The situation would be more understandable as we let the optical size become very small ($\rho \ll 1$) so that the theory need only retain the first term which can give rise to an anisotropy in the internal field.

In this limit the electromagnetic analysis is simple, and leads to the factor J of Eq. (2) of the form

$$J = \frac{3}{8} \left[\frac{12\, \rho \varepsilon_r''(\omega)}{|\varepsilon_r(\omega)+2|^2} \right] \left[\frac{2}{15} \rho\, \varepsilon_r''(\omega) \left(\frac{5/4}{|\varepsilon_r(\omega)+\frac{3}{2}|^2} - 1 \right) \right] \tag{13}$$

where $\varepsilon_r''(\omega)$ is the imaginary part of the relative dielectric constant $\varepsilon_r(\omega)$. This expression has been factored so that the first term in brackets is the absorption efficiency Q_a and the second term is the anisotropy factor A_z. It should be noted that two resonances appear in Eq. (13). The first is a resonance at $\varepsilon_r = -2$, in the absorption, and the second is a resonance in the anisotropy factor at $\varepsilon_r = -3/2$. The first of these is the resonance associated with the surface dipole mode of the sphere, and the second belongs to the surface electric quadrupole mode. A curious result of Eq. (13) is that when the surface quadrupole mode is excited, J may change sign and become positive, so that the radiometric force is reversed. This effect, which is caused by the interference between the quadrupole and dipole modes, should occur in Ag at 347 nm, and would render the field at the back of the particle to be considerably larger than that at the front! A typical distribution of these two fields, just below the dipole resonance, is shown schematically in Fig. 5, with the relative phase of the two modes in front and back of the particle as shown. A comparison of the results obtained from Eq. (13) with a full Mie calculation (i.e., including all modes) shows that Eq. (13) is valid within 15% for silver particles of radii less than and up to about 100 Å. Thus, the major portion of the effect is due to these two lowest modes alone.

The field reversal predicted in these very small particles was so surprising that we extended the Mie calculations to an 850 Å radius, where experiments are tractable. Fig. 6 shows the results of these calculations at a nearby wavelength of 355 nm. This figure is a topological plot of the square modulus of the internal field within an equatorial slice of the particle. We observe directly the reversal in the

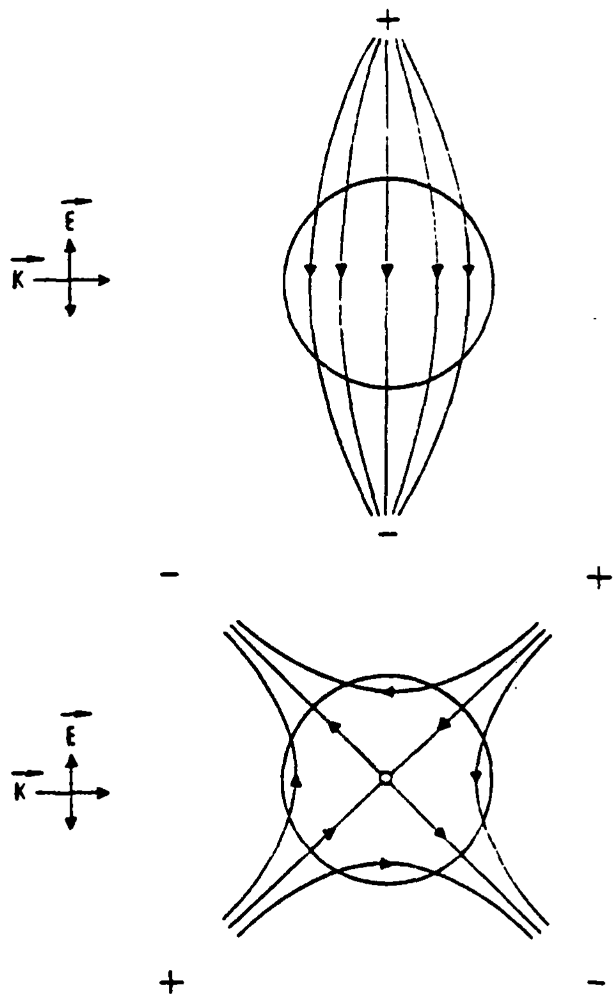


Fig. 5 Dipole and quadrupole field distributions in a small sphere close to the quadrupole resonance. The fields have the same direction at the back of the particle, and opposite directions at the front, producing anisotropy.

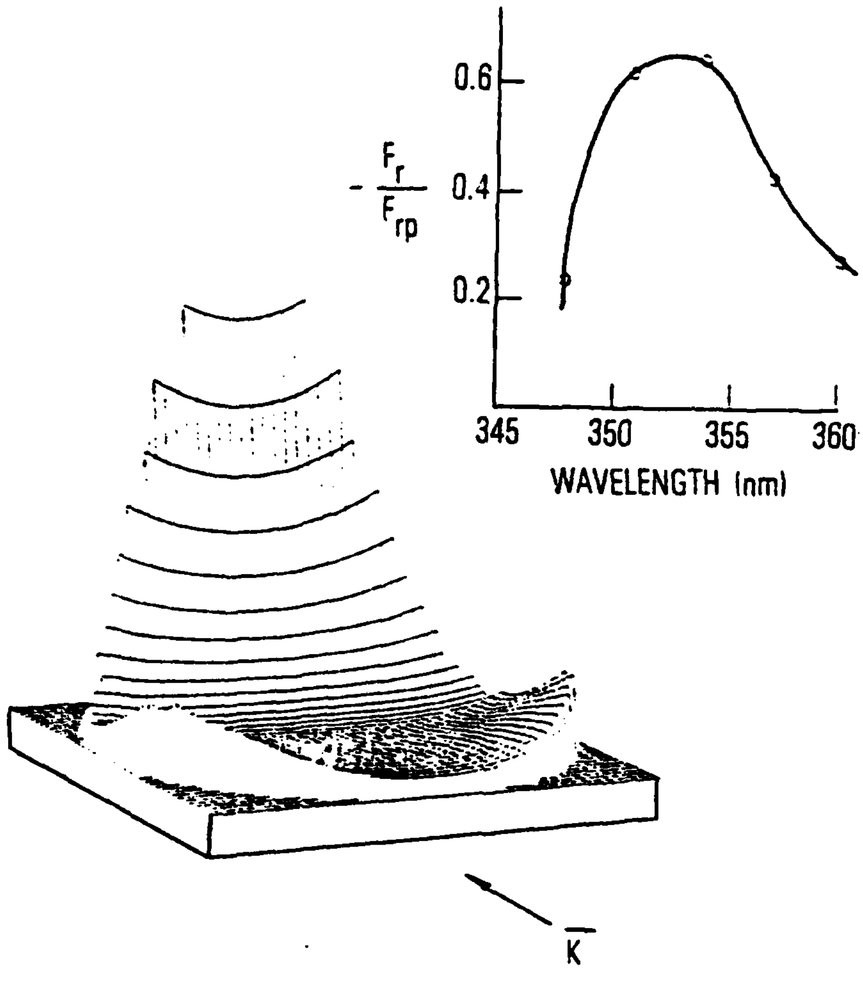


Fig. 6 Front to back distribution of the internal field intensity in a silver particle of 850Å. The inset gives the ratio of the radiometric force F_r to the photon pressure force F_{rp}.

internal sources of heat generation proportional to the field intensity. The inset in this figure gives the ratio of the radiometric force F_r to the radiation pressure force F_{rp} (which is always forward), assuming the bulk thermal conductivity of silver. This ratio is just below unity, and suggests that if the thermal conductivity in small silver particles is somewhat reduced, the net force on silver particles can also reverse.

The effect of competing resonances, and radiometric force reversal, is not expected to exist only for surface plasmons in metals. A similar effect should occur in ionic solids when associated with surface phonons (e.g., in SiC, for which $\varepsilon_r = -3/2$ at 10.63 microns). In fact, while the large thermal conductivity of a metal such as Ag limits the absolute radiometric force so that photon pressure is dominant, the effective resonance in SiC at infrared wavelengths is sufficiently large so that the overall force is reversed, even assuming unchanged bulk thermal conductivity. It would be interesting to look at semiconductors such as InSb in which the thermal conductivity is less than 1/20 that of SiC, and surface plasmons can be steered in frequency by appropriate doping.

A paper on the influence on photophoresis of the interference of surface modes has been published.[52]

We saw above that the anisotropy effect summarized by (13) is due to the presence of a quadrupole mode in addition to the expected dipole mode, and that the anisotropy is most pronounced at or near the quadrupole resonance for the small particle.

In B,(4), we have shown that a simple formulation may be employed to determine the conditions for resonance for spherical particles. This formulation involves spherical transmission line theory, which is a rigorous technique in electromagnetics; a summary that is relevant to our needs here is presented in the above-mentioned section. The simplicity in the formulation is achieved because the method yields the partial information we seek, i.e., the resonances of this class of spherical modes, in a direct manner, without unnecessary appendages.

The general rigorous equation for the resonances of spherical E modes is given earlier as equation (8). It is used in B,(4), to derive directly the well-known result for the lowest, or dipole, mode in the limit of small particle size, showing that the result is, as given by (11),

$$\varepsilon_r = -2$$

Very rarely has anyone been concerned, in the small size limit, with other than this lowest mode, which corresponds to mode number n = 1. However, the condition on the value of ε_r for the resonances of higher modes in the small size limit can also be readily obtained from the general relation (8).

In the small size limit (kR << 1), the expressions for the spherical Hankel and Bessel functions become

$$\hat{H}_n^{(2)}(kR) = \frac{j(n - \frac{1}{2})!\ 2^n}{\sqrt{\pi}\ (kR)^n} \tag{14}$$

$$\hat{J}_n (kR \sqrt{\varepsilon_r}) = \frac{(kR\sqrt{\varepsilon_r})^{n+1} \sqrt{\pi}}{(n+\frac{1}{2})!\ 2^{n+1}} \tag{15}$$

When the derivatives of (14) and (15) are taken with respect to their arguments, and substituted, together with (14) and (15) into relation (8), one obtains

$$-\frac{kR}{n} = \sqrt{\varepsilon_r}\ \frac{k\sqrt{\varepsilon_r}\ R}{n+1}$$

or

$$\varepsilon_r = -\frac{n+1}{n} \tag{16}$$

Relation (16) is a generalization for arbitrary mode n of the familiar relation (11) for the lowest mode, n=1. For n=2, corresponding to the quadrupole mode (which is of importance in the explanation of the phenomenon discussed above, involving small particles which can move backwards), we see that $\varepsilon_r = -3/2$ at the resonance. We may also note that for high values of n the resonances all approach the limit $\varepsilon_r = -1$. In addition, since the value of ε_r usually varies rapidly with frequency, the resonances for the n=1 and n=2 modes should occur very close to each other in frequency. In fact, even for silver particles, for which the intrinsic loss is small, it is found that the resonance curves for the two lowest modes overlap each other significantly.

There is an extremely interesting implication in the above results from a conceptual standpoint. It has always been assumed that, in the limit of very small particle size, the electromagnetic scattering is of the Rayleigh type, i.e., the scattered field distribution is that corresponding to a dipole mode. Even in the vicinity of a plasmon resonance for these small particles, the field is still assumed to be of dipole shape, although its amplitude is greatly enhanced by the resonance. We see above, however, that at a quadrupole resonance the scattered field will not be of the Rayleigh type, but instead of quadrupole form or a combination of dipole and quadrupole resonances. Thus, the scattered field due to small particles may be of Rayleigh form over most of the frequency range, and under most conditions, but the exceptions can yield new and interesting physics.

To complement our work on photophoresis we have developed a new spectroscopy for the independent measurement of the absorption efficiency Q_a, which enters as a factor in the expression for J in Eq. (2). This method, referred to as Structure Resonance Modulation Spectroscopy (SRMS) allows one for the first time to make direct measurements of absorption efficiency on a single isolated microparticle, in the infrared. SMRS utilizes the properties of the natural electromagnetic resonances of a sphere and provides a means for constructing broadband absorption spectra at low intensities.

The resonances of a nonabsorbing sphere of size about 1 micron are seen as extremely narrow spikes in the scattered light excitation spectrum when the particle is subject to irradiation in the visible.[53] These resonances are so narrow, in fact, that a fractional change of size as small as 10^{-4} can completely shift a narrow resonance past a laser of constant frequency.[49] For a particle of 5 micron diameter this represents a change in radius of 2.5 Å. Furthermore, a small fraction of this change can be detected if the radius is systematically varied and the variation of scattered light is detected with phase-sensitivity. In practice, we have detected changes in average radius below 0.01 Å. If the size change is caused by thermal expansion due to the absorption of, say IR, then an IR absorption spectrum may be constructed by observing the modulation of scattered visible light.

Figure 7 summarizes schematically the elements of the required spectrometer. The particle is held stationary in a quadrupole trap (not shown), and the IR is provided by a GLOBAR plus optical wedge monochromator combination. The visible probe dye laser is positioned in wavelength near a structure resonance, and its elastically scattered light is viewed at 90°.

Our first experiment has been carried out on a 5.4 micron diameter drop of a solution of $(NH_4)_2SO_4$ held in size equilibrium with water vapor. The SRMS IR spectrum of this particle is presented in Fig. 8. The well-known line seen in the figure is the absorption of the $SO_4^{=}$ ion. The solid curve is a Mie calculation of Q_a for a 5.4 micron particle of a composition consistent with the vapor pressure equilibrium of the solution. The agreement between the curve and the experimental points is very good, and verifies that the SRMS spectrum is the spectrum of the absorption efficiency of Q_a. It is important to note that this particular sample represents the smallest solution mass (90 picograms) for which an IR spectrum has ever been taken.

A paper on Structure Resonance Modulation Spectroscopy has been published.[54]

(6) Recent Progress

(a) New Electrodynamic Levitator

For the Photophoretic Spectroscopy studies discussed earlier, we employed a modified Millikan chamber (Fig. 2) to levitate a single particle in a stable fashion. When the diameter of the particle is about one micron or larger, as it is for those studies, that mechanism for levitation has been very successful. In the future, however, one of our major interests is in the confinement of particles with diameters down to 100 Å or so. At such a size, the Millikan chamber shown in Fig. 2 is overwhelmed by Brownian motion, so that another scheme must be developed.

With that goal in mind, a new electrodynamic levitator has been constructed during this period that should have a confinement energy orders of magnitude above kT. It is based on a quadrupolar field geometry, and it has already been used very successfully for the work on Structure Resonance Modulation spectroscopy. Although this new

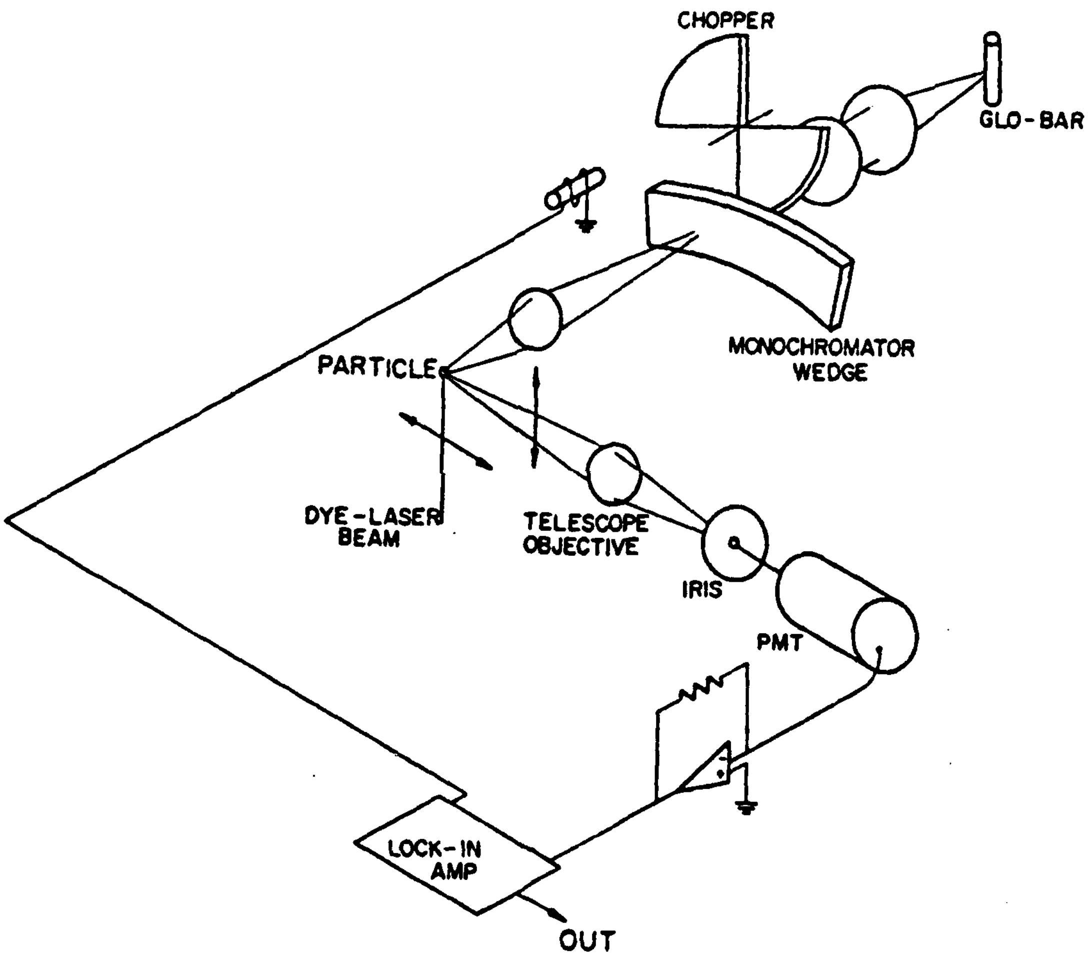


Fig. 7 Schematic arrangement of a Structure Resonance Modulation Spectrometer.

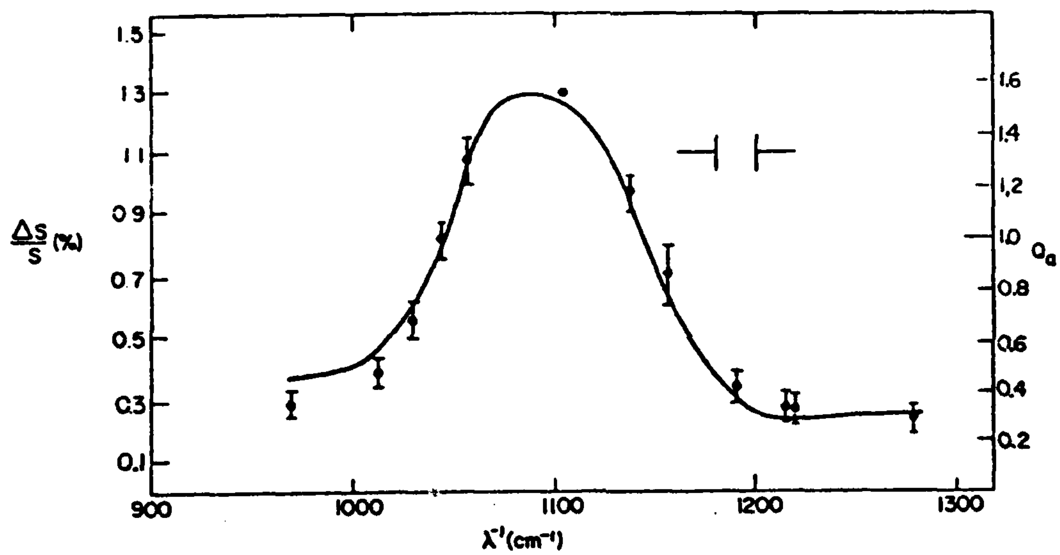


Fig. 8 SMRS result for a $(NH_4)_2SO_4$--water drop. Experimental points: relative change in visible scattered light intensity $\Delta S/S$; full curve: theory of Q_a for this 5.4 micron particle. The horizontal spacer in the upper right gives the bandwidth of the incident IR.

levitator has so far been used only for micron-sized particles, it is expected that with minor additional considerations in the design it can also work well for particles of 100 Å size. We will make the necessary modifications as part of the proposed research program.

Although originally introduced by Paul as a mass filter,[57] the quadrupolar concept has been used in atomic physics for trapping single atomic and molecular ions for extended periods.[58] In addition to adopting Paul's filter, we have automated it as in the case of the Millikan capacitor. We can now anticipate the measurement of particle mass by electron stepping[34] down to 10^{-18} gm.

The electrodynamic balance is shown in Fig. 9. It consists of an electrode configuration that produces a quadrupolar field geometry. When the electrodes are driven with an ac voltage, small charged particles remain in this device because of the time-averaged upward force which counter-balances gravity. At the center of the device the ac produced field vanishes and the dc field produced by the position's sensitive detector balances the particle. To our knowledge, this is the first automated electrodynamic balance.

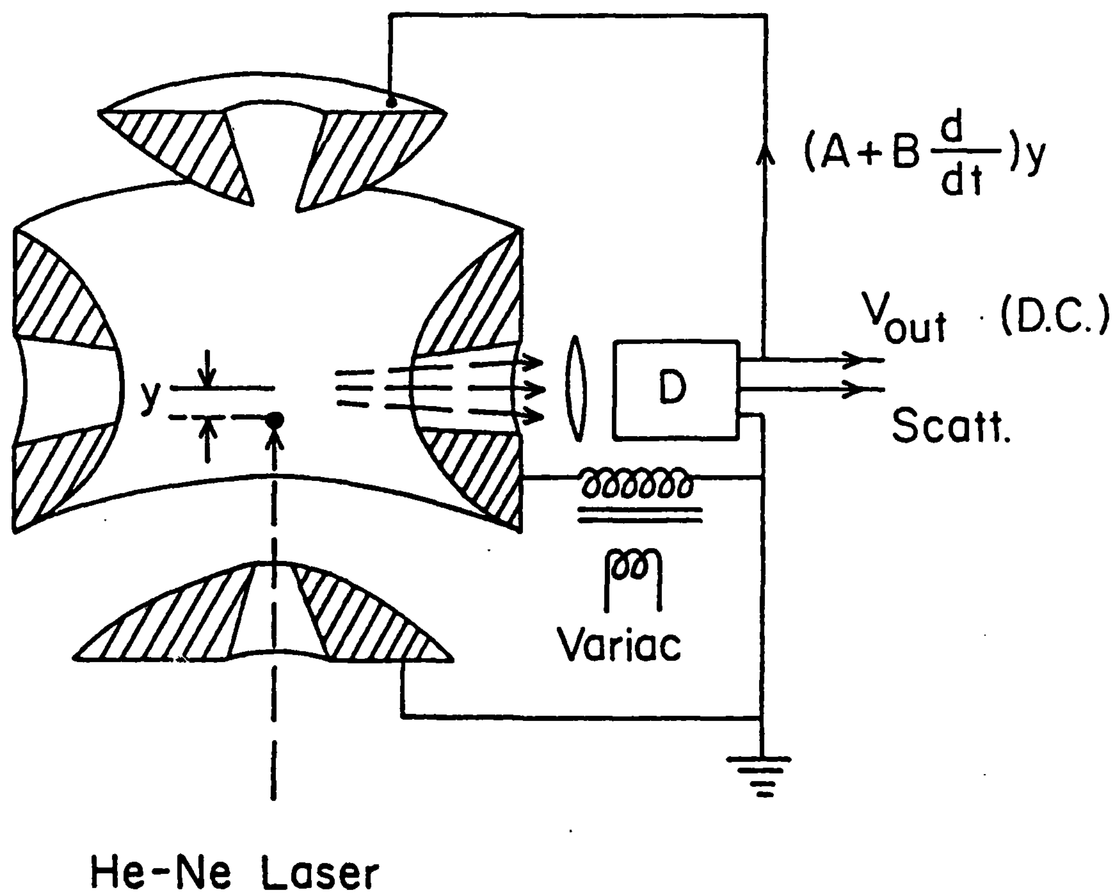


Fig. 9 Schematic of automated electrodynamic levitating balance. D is a position-sensitive detector, providing a voltage feedback dependent on the vertical displacement y and its time derivative. The output of D gives the voltage (force) required to stabilize the particle.

(b) The Determination of the Optical Response of a Single Microstructure About 100 Å in Size

Most of what is known about the interaction of electromagnetic radiation with a micron-sized object has utilized the bulk dielectric function $\varepsilon(\omega)$. It is difficult to understand how this practice can be carried over to objects 100 Å in size. After all, the dielectric function in metals, the Drude equation, is based on diffusive transport, while the mean-free path of an electron in most metals at room temperature is >100 Å. In other words, the classical theory of absorption and scattering should break down for structures of 100 Å in size. Figure 10 shows the results of a calculation by Professor Leung of the elastic scattering for a 50 Å metallic particle, in comparison with the curve based on a Drude dielectric response. The resonances due to Fermi wavelength effects are apparent and give large enough effects so that they should be observable in a single particle experiment. Calculations of absorption[51] show even more pronounced effects, with the far infrared absorption being larger by a factor of 10^2 than the classical calculation.

(c) Resonances on Tapered Fibers

The use of tapered fibers provides us with a continuously variable microstructure in the sense that the wavelength for a given circumferential mode varies linearly with displacement along the fiber axis. Preliminary experiments have shown that tapered fibers produced in our laboratory have high Q resonances. These resonances are evident in the elastic scattering excitation spectrum measured on a short section (about 10 microns long) of a fiber having a 5 micron radius (Fig. 11). This paves the way for investigating whether or not the rate of spontaneous emission by an excited state-structure system differs from that by the excited state alone.

C. Optical Interactions at Surfaces with Controlled Roughness

One of the nearly universal conditions reported for the enhancement of Raman scattering at silver surfaces is that the surface must exhibit roughness. The common interpretation of this requirement is that roughness provides, at least statistically, the local curvature of the surface that produces the conditions for the buildup of optical fields. This buildup, of course, manifests itself in other optical interactions, as well, especially nonlinear ones that depend on higher powers of the incident and reradiated intensities such as harmonic generation, two or more photon transitions, for example. However, there is considerable uncertainty as to the actual extent of the roughness required. Most film samples are used to form a structure of metallic clusters over several hundred Å in dimension,[20] while a minimum of 100 to 150 Å rough skin on a smooth metal surface has been found necessary in order to produce enhancement in at least one set of experiments.[21] Generally, the nature of the roughness, especially the character of the minimum roughness that is required to see field enhancements, remains unclear. A better determination of the conditions for the onset of field enhancement will contribute to elucidating the extent to which these enhancements result from classical electromagnetic local field buildups alone, and to which other, probably more atomistic-scale, mechanisms also play a role.

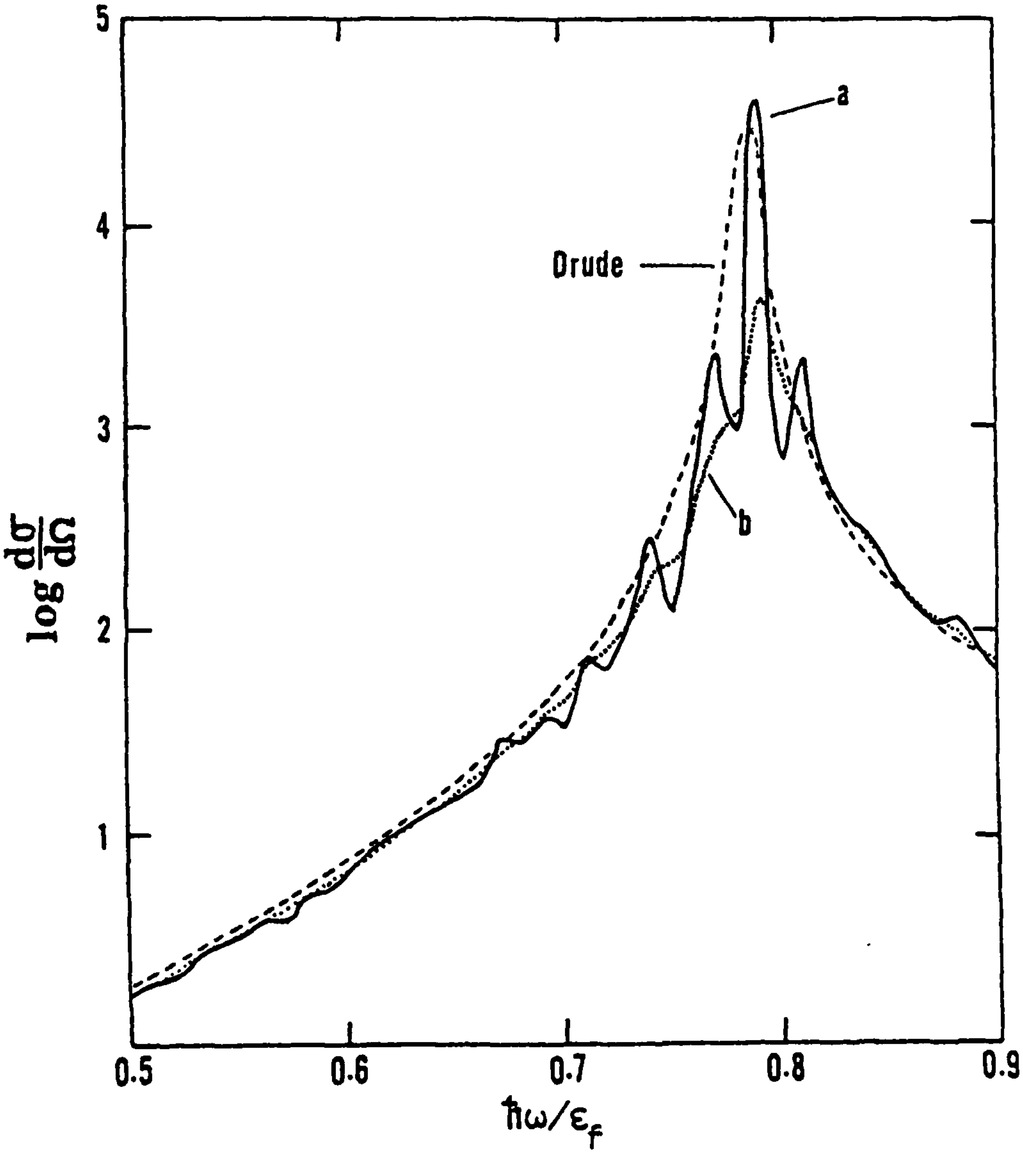


Fig. 10 Elastic scattering from a 50 Å metallic sphere as a function of frequency. The electron mean free path is ℓ = 700 Å for curve a and ℓ = 160 Å for curve b.

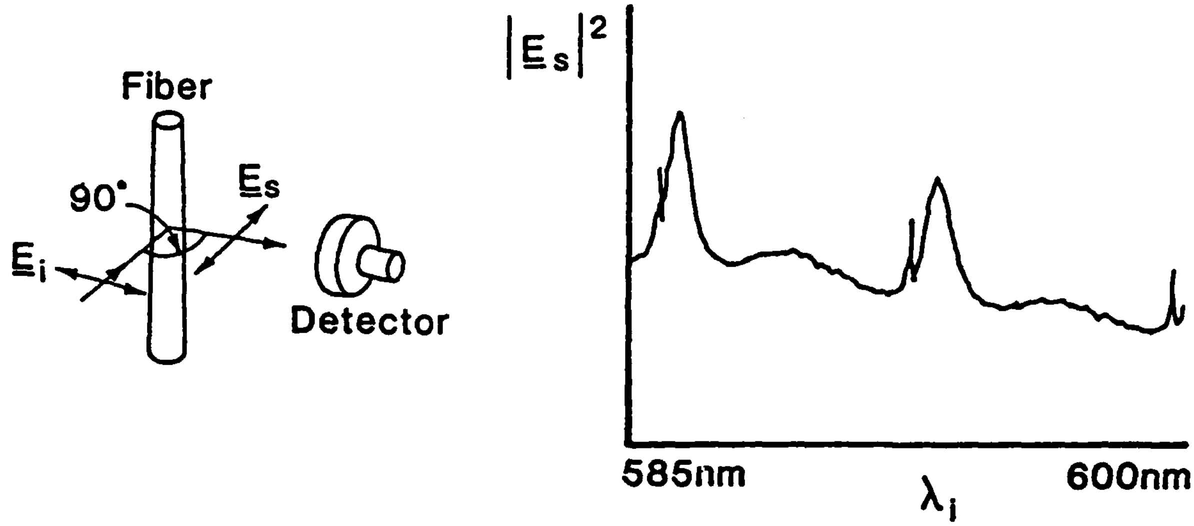


Fig. 11 Sharp scattering resonances on tapered fibers. Experimental arrangement and results. The scattered light is viewed 90° from the incident direction.

We are examining this question using an approach that emphasizes correlating the optical responses of a metal surface, most importantly its diffuse scattering and its second harmonic generation, with surface roughness, as this roughness is changed systematically. In another work unit of this report (SS4-2), we have carried out a program for characterizing the roughness imposed on an initially very flat and smooth surface by an analysis of how internal electrons scatter from such a surface. In particular, by varying the sample temperature, this roughness can be altered practically at will. Here we correlate the information obtained by such internal electron scattering with the external optical scattering behavior. The first method of measurement is most sensitive for very thin rough layers, but overlaps the range in which the onset of optical field buildup is expected, and it is, of course, ideal for probing smoother surfaces where atomic scale roughness must predominate.

In the newly beginning project on the measurement of the optical properties of metallic surfaces showing a variety of microscopic surface structure, the primary emphasis had been on the development of a portable vacuum system that allows the good preparation and characterization of diverse surfaces, and that can then be brought up to light sources in different frequency ranges located in different experimental setups, so that the sample surface conditions can be maintained. While the system maintains a good vacuum under equilibrium conditions, we have found that during evaporation the background pressure due to degassing of the metallic source rises to levels that make fast deposition difficult. The surfaces produced under such conditions do not have the purity and the degree of perfection that we used for a definitive measurement. An auxiliary system to handle the gas load during the deposition time is being added.

Most recently the optical part of the experiment on second harmonic properties of flat metallic surfaces has been made operative. In order to initially avoid the loss mechanisms induced by exciting surface plasmons, we have chosen to use an incident wavelength near one micron, and to detect the signal appearing in the violet. Use of synchronous photon counting techniques and a pulsed diode laser has resulted in good sensitivity to detect the second harmonic on flat surfaces, where the interaction is, of course, much smaller. We are in the process of determining the absolute value of the conversion constant, below that found on rough surfaces, as a preliminary for studying the roughness dependence of the conversion on surfaces with fixed roughness prepared and kept under vacuum. The results so far show full isotropy for polycrystalline surfaces, but indicate a definite angular structure for single crystal 111 surfaces. Their full origin remains to be determined, but is probably connected with the loss of centrosymmetry in the first few atomic layers at and below the crystal surface.

D. Electromagnetic Theory of Plasmon Resonances on Small Non-spheroidal Metal Particles, Including the Effects of Array and Other Environments

In this portion of the program, we assume that the enhancement which involves metallic particles is due solely to macroscopic electromagnetic causes, and we pursue certain analyses that can furnish significant information and insight regarding the enhancement processes, particularly those features such as departure from spheroidal shape and effects due to mutual interaction.

There are two principal effects, both purely electromagnetic, that combine to produce the remarkable values of enhancement, namely,

(a) surface plasmon resonances on the metal particles, which are dependent on size and shape, and

(b) especially strong concentrations of electric field, which are effective when the excited molecule is properly located.

Effect (b) follows from the specific shape of the particle and also the particular resonance involved. This effect has been termed the "lightning rod" effect by Gersten and Nitzan,[22] and it indicates that electric fields are greatly increased in the neighborhood of regions of strong curvature. Gersten and Nitzan treat only prolate spheroids, but for maximum enhancement they place their active molecules on the axis just above the pointiest part of the prolate spheroid. For a 5 to 1 axial ratio for a prolate spheroid, they calculate an enhancement as high as 10^{11}, in the region of red light, for an optimally-located active molecule. If the metallic particle could be shaped with a very sharp edge or corner (which would involve a departure from the spheroidal shape), the enhancement could be even greater since the electric field becomes divergent at an ideally-sharp edge or corner if the electric field has a component perpendicular to that edge or corner. However, only a few well-located molecules will be treated that well; molecules in average locations will participate in lower enhancements, and will not benefit from the "lightning rod" effect.

Almost all of the active molecules, on the other hand, will benefit from effect (a), since the plasmon resonances increase the fields everywhere on the metallic particle. This portion of our program will deal primarily with effect (a), which contains most of the interesting physics in addition to the unsolved problems.

As mentioned earlier, the surface enhancement of Raman scattering was discovered on rough surfaces of metals, particularly silver. These rough surfaces were mathematically modeled by spheres, or oblate or prolate spheroids, and some agreement with experiment was found. It was shown that the near fields peaked at wavelengths for which plasmon resonances were achieved. These resonances were shape-dependent, and corresponded to an equivalent dielectric constant for the metal that ranged from near -1 all the way to -10 or more, in the region of red light, for extreme prolate spheroids.

The theories available are only for spheres or for oblate or prolate spheroids, although in general the theories for those shapes are very good. The most complete analysis for spheres was presented by Kerker, Wang and Chew,[23] and that for prolate spheroids was given by Gersten and Nitzan.[22] However, these theories are for isolated particles, and in air (or a homogeneous liquid), not on a silver surface or on a dielectric substrate which can act as a support. In addition, interaction with neighboring particles has not been taken into account in those theories, although the problem of collective electron resonances was discussed in connection with SERS by Moscovits[24] and by Burstein, Chen and Lundquist.[25]

More recently, it was realized that rough surfaces were not efficient since, for a given wavelength, only some of the particles participated in the resonance. It was therefore proposed that a periodic structure be employed, and that the silver particles all be made identical to each other, thus sharpening the resonance and concentrating the effect. Towards this end, Liao et al.[26] presented such an experiment; the structure was fabricated at MIT and measured at the Bell Laboratories. The structure consisted of a two-dimensional array of dielectric (SiO_2) support posts with identical silver blobs deposited on the top of each post; each silver blob resembled somewhat a prolate spheroid with a 3 to 1 axial ratio. Experimentally, they found a resonance-shaped response with frequency, which changed appropriately as the surrounding medium was changed, but the agreement with theory for a 3 to 1 prolate spheroid was mediocre at best. The authors point out that the agreement was not that good because of three limitations in the available theories:

(1) they do not permit departures from the spheroidal shape,

(2) they do not account for the presence of the dielectric support posts, and

(3) they neglect mutual interaction effects between neighbors.

It is clear, therefore, that there is a need for a broader theory that will take these effects into account. In this portion of this work unit, we address the limitations listed as (1) to (3) above.

The general periodic array structure that we are in the process of analyzing is shown in Figure 12. It is a two-dimensional rectangular array of dielectric support posts, placed on a dielectric substrate which can be of similar or different material, and with silver deposits on each of the support posts. We are selective here in choosing the cross sections of the dielectric support posts and the silver deposits to be rectangular, since they are then compatible with the rectangular nature of the array. The solution for the structure then becomes tractable. In addition, the rectangular shape is a likely result of most of the actual deposition processes.

We analyze the structure in Fig. 12 by choosing a unit cell surrounding a typical support post plus silver deposit. The vertical walls of the unit cell are so-called "phase-shift walls," the relative phase shift between them being dependent on the angle of the incident plane wave and their nature dependent on the polarization of the incident wave. This approach automatically takes mutual interactions into account. The rectangular shapes involved permit us to handle the solution in an approximate but accurate way, by taking advantage of techniques recently developed by us in connection with studies of open dielectric waveguides.

The basic phrasing of the problem in terms of unit cells and phase-shift walls employs techniques in phased-array antennas which we ourselves have pioneered.[27,28] The application to this problem involves certain important differences from the direct phased array experience, but the basic approach is similar.

Before tackling the periodic array of Fig. 12, which involves several complications simultaneously, we are analyzing several simpler configurations to develop our understanding more systematically. First, we have been treating the isolated parallelepiped shape, corresponding to the silver deposit of Fig. 12, by itself. That analysis determines the resonances for that structure as a function of size and shape. The parallelepiped geometry permits a greater variety of shapes than the spheroidal geometry does. Thus, these analyses of the isolated structure should yield insight into where the wavelength resonances appear when deviations from the spheroidal shape occur. The method of analysis employed treats the particle as a waveguide with open end terminations, and takes them into account. Use has been made of solutions for surface waves on thin overdense plasma layers originally derived in another context some years ago by Oliner and Tamir.[29,30]

Of all the possible resonant field configurations corresponding to permissible symmetries, only one yields a condition for resonance which, in the limit of dimensions small compared to wavelength, is independent of frequency and is a function only of the ratios of dimensions. This condition results in a cubic relation for the relative dielectric constant, which of course must be negative real when the metal is taken to be lossless. We have derived the expression and are examining its numerical consequences.

At the same time, we have begun to examine a two-dimensional rectangular array of such parallelepipeds in space to determine the changes due only to mutual interactions. We employ the unit cell approach and phase-shift walls, but in a simpler context. The mutual

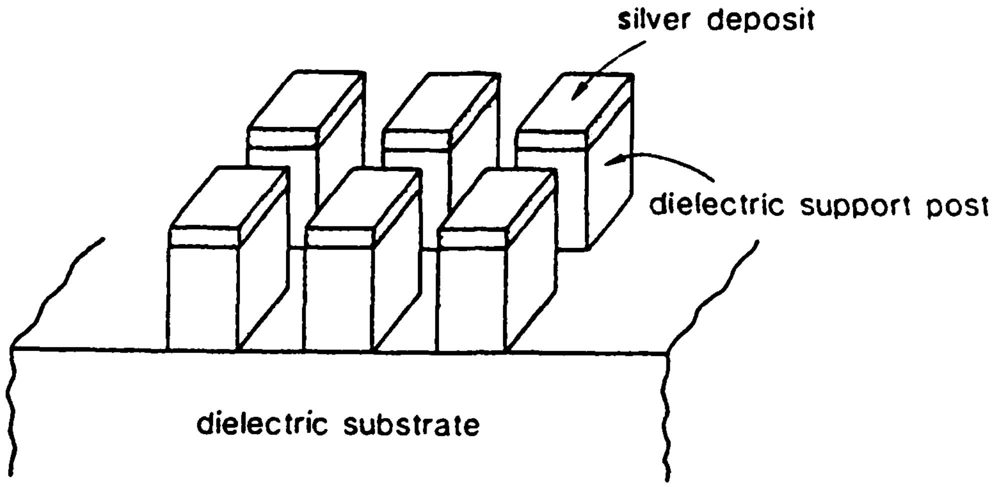


Fig. 12 Two-dimensional rectangular array of silver deposits on dielectric support posts located on a dielectric substrate. This periodic structure is intended to closely model recently-proposed array structures for concentrating the enhancement effects.

interactions are taken into account automatically by this approach, and we aim to determine the shift due to the array environment as a function of size, shape, and array spacings. We would see if we could extrapolate the results to predict the shift due to interactions if spheres or spheroids were involved.

In addition, a very interesting but simple result has already been found in connection with the metal particle of rectangular shape placed (or grown) on a flat metal surface.

A recent paper[46] by Wood presents experimental data in connection with an added layer of rough silver particles on a smooth silver flat surface. He finds that very thin rough films do not exhibit SERS effects, whereas much thicker films of that type do. He concludes that his experiments disprove the adatom theory of the effect. He observes, in fact, that Raman enhancement occurs only when the roughness features exceed about 150 Å or so. By simple considerations in electromagnetics we can show why the enhancement effects should not appear when the roughness features are too small.

We take as our model a small metal rectangular block placed on a metal plane. A metal half space exists under the metal plane. Similarly, a section of a half space is also present under the metal block, the difference being that that portion of metal half space is slightly higher than the remainder, and that a step junction on each side separates the two regions. Next, we remember that the metal acts as an overdense plasma, so that a surface plasmon is present at the air-metal interfaces when light is shone on the structure. The surface plasmon decays into the air region above and into the metal region below. If the plasmon is tightly bound to the surface, so that the step junction is large relative to the plasmon's transverse (vertical) extent, the plasmon is trapped in the block region and the Q of the resonance (when it occurs) will be very high. The Raman enhancement will then be significant. If the reverse is true, that is, the step junction is electrically small, the plasmon will simply leak past the step junction, so that the Q of the resonance will be very low and the enhancement negligible.

The decay into the metal region may be characterized by a skin depth δ, given by

$$\delta = \frac{|\varepsilon(\lambda_0)|-1^{\;1/2}}{|\varepsilon(\lambda_0)|} \cdot \frac{\lambda_0}{2\pi} \qquad (17)$$

which represents the vertical distance in which the field has decayed to 1/e of its value at the surface. If the value of δ is large, the presence of the step junction between the block and the flat surface will not be felt unless it is also large.

Calculations for silver as a function of wavelength λ_0 are summarized in Table 1. Interestingly, the skin depth values are almost the same over the whole visible spectrum, being of the order of 250 Å or so for 490 nm light, which Wood[46] used. For step sizes (which correspond approximately to roughness features) much smaller than δ we would not expect significant effects. The threshold for enhancement would be expected to occur for step sizes somewhat smaller than δ, and 150 Å, which Wood found for the threshold, fits well into the predictions of this simple model.

Table 1.

λ_0	$\varepsilon(\lambda_0)$	δ
350 nm	- 2.0	280 Å
490 nm	- 9.1	244 Å
620 nm	-17	232 Å

E. Quantum Mechanical Aspects of Local Electromagnetic Fields Close to Metal Surfaces and Small Structures

Although most of the information available on strong local enhancements has been interpreted on the basis of the Mie theory of the scattering of electromagnetic radiation by spheroidal particles, there are some data that do not fit within this framework.[3] Other effects are also at work. Some of these, such as the departures from spheroidal shapes and mutual interaction between neighboring scatterers, are addressed in Section D above. Others are related to the fact that on the scale of molecular dimensions, and similar distances of the molecule above the surface, quantum mechanical effects cannot be ignored. For example, Gersten and Nitzan, in commenting on the image mechanism for enhancement, state that

> "its true nature has to be investigated with the quantum mechanical nature of the surface, the molecule, and their interaction taken into account."[22]

Similarly, Bergman et al question the use of bulk dielectric constants for very small particles where

> "quantum mechanical effects may become important."[31]

Here we have been concerned with some of the quantum mechanical effects that can enhance the local fields near a surface to values above those calculated from purely classical electromagnetic theory, along the lines explored by Feibelman for flat metal surfaces.[9] These effects will be important components of the description of the resonances associated with very small structures and will, for example, be essential in the interpretation of surface photoemission spectra from such structures. As already implied in the quotes above, these considerations will introduce another scale of smallness into the discussion, namely the range of quantum mechanical perturbations due to the existence of boundaries, and the distance below which classical modeling breaks down. This scale will in general be superimposed on the scale set by the wavelength of the electromagnetic radiation, but especially for very small structures, which are classically too thin to show interference effects, the quantum mechanical effects may make the dominant contribution to enhancements and other anomalies.

The effects which we consider are most clearly seen through a straightforward discussion based on an idealized model of a metal. Consider a free electron gas, confined to a volume by a uniform potential which represents the smeared out positive ionic charge (the jellium approximation). The low energy electronic wave functions inside the volume are approximately described as standing waves, which decay exponentially outside the volume. Due to the Pauli exclusion principle electrons will fill up the lowest energy states up to the fermi surface. The resulting charge density will show a large decrease at the boundaries of the volume, and superimposed on this will be Friedel oscillations[32] with a wave vector that corresponds to that of the fermi surface. Thus there will be rapid variations in the charge density near a surface, which contain a significant quantum mechanical contribution. These rapid variations have an important impact on the properties of applied electromagnetic fields, such that the amplitude and phase of an electromagnetic wave will change dramatically in the surface region. Calculations by Feibelman[9] on a semi-infinite solid have shown that the local fields are enhanced by the rapid variations in charge density, by as much as a factor of ten within the outermost layers of the metal surface, and therefore all photon-surface interactions within this region cannot possibly be treated using classically deduced electromagnetic field amplitudes alone.

In particular, these non-classical field enhancements imply that photoemission arising from the immediate surface layer of a metal is completely dominated by them. The theory predicts a large increase above classical estimates of the photoelectron yield, a different frequency dependence of this yield, as well as substantial changes in the emitted photoelectron's angular distribution because of the rapid changes in the dielectric constant of the solid surface with frequency. Recently, there has been convincing experimental evidence that some of these predictions are correct,[33] so that for the first time there exists a quantitative explanation for the surface photoelectric effect.

The general problem of quantum effects on the electromagnetic response of small metallic particles has been approached in terms of formulations for model geometries. The first geometry that has been considered is that of a thin slab. This requires the calculation of the eigenfunctions and eigenvalues for the wave functions of the interacting electrons within two parallel boundaries. The eigenfunctions are translationally invariant parallel to the slab's surfaces. The main difficulty is therefore associated with the variations of the wave functions in the direction perpendicular to the two surfaces. For a thin slab, the finite spacing between quantum levels makes the calculation of the wave functions remarkably simple. If the Fermi energy is μ, one only needs to consider a finite number N of bound wave functions, with N given by

$$N \sim \sqrt{2m\mu}\, L/h\pi .$$

The higher energy continuum states that may be needed in addition are reasonably approximated by phase shifted plane waves.

This approach to calculating the needed wave functions bears some similarity to that used recently by Wood and Ashcroft,[51] although there are some crucial differences. They have included neither the electron-electron interactions nor the finite work function of the slab. Both these features are important in our problem, since they tend to smooth out the variations of the electron density. Since, as Feibelman has shown, it is the gradient of the electron density which controls the quantum enhancement of the electromagnetic fields at surfaces, this smoothing out of the density variations must be known.

The eigenfunctions described above are being used to calculate the electromagnetic response function $\nabla(\underline{r},\underline{r}',\omega)$. The resulting conductivity separates into a local and a non-local part. The local part is given by a Drude-like formula that depends on the local electron density. This density has Friedel oscillations centered about the surfaces of the slab. These are the local variations in the conductivity that Feibelman found to be important in determining the enhancement of the electromagnetic fields at a single plane surface. The non-local response also shows oscillatory behavior. This contribution is not included in the Ashcroft calculation.

We have now obtained the dielectric response, of the thin jellium slab, within the time dependent local density functional approximation. The wave functions were obtained, as by perturbation theory of the starting basis set obtained from the solutions of a finite square well potential. The difference between this potential and the full density functional potential was treated to first order in perturbation theory. The reasonableness of this procedure was assured by comparisons of the density with that obtained by Lang and Kohn for the semi-infinite jellium model of a single surface.

The comparison shows that the perturbation theory for the slab has a slower variation in the electron density near the surface than in the Lang and Kahn scheme. The Friedel oscillation occurs deeper within the slab and the tail of the electron density in the vacuum is larger.

The calculation of the electromagnetic field passing through those slabs has also been carried out. The integral equations for the vector potential were solved for frequencies much lower than the threshold for photoemission. For such processes, only the bound state warefunctions contribute to the electromagnetic response. In this case only a finite number of wavefunctions need be incorporated in the basis state. One of the quantities that is quite simple to evaluate here is the reflectance. This has been studied as a function of the frequency of the incident field. Several resonances are present, below and above the plasmon frequency. The sharpness of these resonances depends crucially on the assumed electron mean free path. The nature of this dependence is more critical than one might expect. A mean free path of even as much as four times the slab thickness can almost completely wash out the sharpness of the resonances. This sensitivity has to be studied further.

The quantum effects on the electromagnetic response of small metal particles are best characterized in the non-local conductivity.

$$\sigma (\vec{r}, \vec{r}; \omega)$$

Systems that are small enough to exhibit quantum size effects have dimensions that are typically no larger than a few hundred Ångstroms. For systems this small, depolarizing fields have an important role in quenching some of the quantum mechanical enhancements. In particular, at very low frequencies, quantum theories have suggested that the infrared absorption should be enhanced over and above the absorption expected on the basis of classical theories. Depolarization fields do, in fact, tend to suppress this expected enhancement. Feibelman[59] has recently published an article covering essentially the same problem, and with the same conclusions. As a consequence, we have now halted further work on this portion of the program.

A particularly interesting set of systems would be those in which quantum size effects may exist in a macroscopic system. A specific example could be random quasi-one-dimensional systems, such as K_2 Pt $(CN_4)_3$ (H_2O) in which the Pt ions form one-dimensional chains. The electronic properties of these systems are dominated by the Pt ions. The dielectric constants of these systems are anomalously large and anisotropic. Typical values of the dielectric constant along the direction of the chains can be as large as 100, and may be varied by intentionally introducing impurities or defects in this system. We have performed a theoretical investigation of the quantum mechanical effects which give rise to this very anomalous infrared response.[55] It appears that the electromagnetic absorption on these systems is composed of many very sharp lines of an entirely quantum origin. The intensity of these is extremely sensitive to vacancies on the Pt sites, very much in analogy with the mean free path sensitivity already mentioned above.

5. REFERENCES

1. T.E. Furtak and J. Reyes, Surf. Sci. 92, 351 (1980).
2. Surface Enhanced Raman Scattering, R.K. Chang and T.E. Furtak, editors, Plenum (1981).
3. A. Schmitt-Ott, P. Schurtenberger and H.C. Siegmann, Phys. Rev. Lett. 45, 1284 (1980).
4. C.K. Chen, A.R.B. deCastro and Y.R. Chen, Phys. Rev. Lett. 46, 145 (1980).
5. A.M. Glass, P.F. Liao, J.G. Bergman and D.H. Olson, Optics Lett. 5, 368 (1980).
6. S.L. McCall, P.M. Platzman and P.A. Wolff, Phys. Lett. 77A, 381 (1980).
7. J.O. Gersten, D.A. Weitz, T.J. Gramila, A.Z. Genack, Phys. Rev. B 22, 4562 (1980).
8. R. Dornhaus, R.E. Brenner, R.K. Chang and I. Chabay, Surf. Sci. 101, 367 (1980).
9. P.J. Feibelman, Phys. Rev. B 12, 1319 (1975).
10. A. Otto, et al., "The adatom model, how important is atomic scale roughness?" in Ref. 2.
11. S. Arnold, Y. Amani and A. Orenstein, Rev. Sci. Instrum. 51, 1202 (1980).
12. S. Arnold, to be published in J. Appl. Phys.
13. R.K. Chang, private communication.
14. A.B. Pluchino, Appl. Opt., 20, 531 (1981).
15. F. Ehrenhaft, Phys. Z. 18, 352 (1917).
16. Y.I. Yalamov, V.B. Kutakov and E.R. Schukin, J. Colloid Interface Sci. 57, 564 (1976).
17. G. Mie, Ann. Phys. 25, 377 (1908).
18. M. Kerker, D. Wang and H. Chew, Appl. Optics 19, 4159 (1980).
19. L. Felsen and N. Marcuvitz, Radiation and Scattering of Waves, Prentice Hall, New Jersey (1973), Secs. 2.5 to 2.7.

20. A. Wokaun, J.G. Bergman, J.P. Heritage, A.M. Glass, P.F. Liao and D.H. Olson, to be published in Phys. Rev. B.

21. T.H. Wood and M.V. Klein, Solid State Commun. 35, 263 (1980).

22. J. Gersten and A. Nitzan, J. Chem. Phys. 73, 3023 (1980).

23. M. Kerker, D.S. Wang and H. Chew, Appl. Optics 19, 4159 (1980).

24. M. Moskovitz, J. Chem. Phys. 69, 4159 (1978).

25. E. Burstein, C.Y. Chen and S. Lundquist, Proceedings of the Joint US-USSR Sympos. on the Theory of Light Scattering by Condensed Matter, Plenum Press, New York (1979).

26. P.F. Liao, J.G. Bergman, D.S. Chemla, A. Wokaun, J. Melngailis, A.M. Hawryluk and N.P. Economou, preprint.

27. S. Edelberg and A.A. Oliner, IRE Trans. Antennas and Propag., AP-8, 286 (1960).

28. A.A. Oliner and R.G. Malech, Chapters 2, 3 and 4 of Microwave Scanning Antennas, Vol. II, "Array Theory and Practice," Ed. R.C. Hansen, Academic Press, New York (1966).

29. A.A. Oliner and T. Tamir, J. Appl. Phys. 33, 231 (1962).

30. T. Tamir and A.A. Oliner, Proc. IEEE 51, 317 (1963).

31. J.G. Bergman, et al, Optics Lett. 6, 33 (1981).

32. J. Friedel, Phil. Mag. 43, 153 (1952).
P. Riseborough and D.C. Mattis "Sum Rules in the Kondo Problem," (unpublished).

33. H.J. Levinson, E.W. Plummer and P.J. Feibelman, Phys. Rev. Lett. 43, 952 (1979).

34. S. Arnold, J. Aerosol Sci. 10, 49 (1979).

35. H. Kuhn, J. Chem. Phys. 53, 101 (1970).

36. P. Chylek, J. Opt. Soc. Am. 66, 285 (1976).

37. S. Arnold and M. Lewittes, "Size Dependence of the Photophoretic Force," J. Appl. Phys., 53, 5314 (1982).

38. S. Arnold, M. Pope and T.K.T. Hsieh, Phys. Stat. Sol. (b) 94, 263 (1979).

39. P.S. Riseborough and D.L. Mills, in Valence Fluctuations in Solids, p. 229, L. Falicov, W. Hanke and M.P. Maple, eds., North Holland (1979).

40. P.S. Riseborough and D.L. Mills, Phys. Rev. B 21, 5338 (1980).

41. P.S. Riseborough, Sol. State Comm. 38, 79 (1980).

42. P.S. Riseborough, p. 225, Ref. 39.

43. N.D. Lang and W. Kohn, Phys. Rev. B 1, 4555 (1970).

44. E. Matijević and D.M. Wilhelung, to be published in J. Colloid and Interface Sci.

45. E. Borgarello, J. Kiwi, E. Pelizzetti, M. Visca and M. Grötzel, Nature 289, 158 (1981).

46. T.H. Wood, Phys. Rev. B, 24, No. 4, 2289 (1981).

47. A.B. Pluchino, Appl. Opt. 23, 103 (1983).

48. S. Arnold and A. Pluchino, to be published.

49. S. Arnold and A. Pluchino, "IR Spectrum of a Single Aerosol Particle by Photothermal Modulations of Structure Resonances," Applied Optics, 21, 4194 (1982).

50. M. Lewittes, S. Arnold and G. Oster, "Radiometric Levitation of Micron Sized Spheres," Appl. Phys. Lett., 40, 455 (1982).

51. D.M. Wood and N.W. Ashcroft, Phys. Rev. B, 25, 6255 (1982).

52. S. Arnold, A.B. Pluchino & K.M. Leung, "Influence of Surface-Mode Enchanced Local Fields on Photophoresis," Phys. Rev. A 29, 654 (1984).

53. A. Ashkin and J.M. Dziedzic, Appl. Opt. 21, 4194 (1982).

54. S. Arnold, M. Neumann, and A.B. Pluchino, "Molecular Spectroscopy of a Single Aerosol Particle," Opt. Lett. 9, 4 (1984).

55. P.S. Riseborough, Z. Phys. B - Condensed Matter 51, 173 (1983).

56. K.M. Leung, Solid State Comm. 50, 449 (1984).

57. W. Paul and M. Raether, Z.f. Phys. 140, 262 (1955).

58. H.G. Dehmelt, IEEE Trans. on Inst. and Meas. IM-31, 83 (1982).

59. P.J. Feibelman and D.R. Hamann, Phys. Rev. B, 29, 6463 (1984).

B. SURFACE STRUCTURAL PROPERTIES OF METALS AND CONDUCTION-ELECTRON SURFACE SCATTERING

Professor H.J. Juretschke

Unit SS4-2

1. OBJECTIVE(S)

To study and characterize the fault structure of the surfaces of nearly perfect thin films of the noble metals, and of very thin irregular overlays deposited on such surfaces at low temperatures in order to understand the mechanisms leading to thermal equilibrium, and the strain distribution in equilibrated surfaces; also to elucidate the very anomalous low temperature behavior of the conductance recently observed by us when the bulk of the thin film is strained by application of a surface charge.

2. APPROACH

To measure internal electron scattering from these surfaces through surface conductivity and its modulation by externally applied surface charge, and then relate the surface scattering to appropriate surface disorder parameters of a new model theory; and to observe internal strains, and their possible low temperature shielding by the strain sensitivity of the electrical conductivity.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

The work during the past year has concentrated on several separate, but related, aspects of the general topic under investigation in this unit.

1) Effective penetration depth of electric fields at metallic surfaces

We have confirmed that the behavior observed at a silver-mica interface, when a thin graphite layer is interposed, also carries over to the silver-vacuum interface, leading to an effective penetration depth of about 15 Å. Attempts to utilize other materials besides graphite to establish a thin layer of poor conducting properties, yet small electrostatic screening length, have had only limited success because of interdiffusion and interface scattering by conduction electrons. Preliminary work using an alumina shows some promise, although the full complexity of a definitive layer arrangement utilizing such a barrier has yet to be tested.

2) Strain sensitivity of the resistivity of silver at low temperatures

An experiment to measure this property on bulk silver has been designed and is being tested out. It shows sufficient sensitivity to provide the needed information for interpreting some of the earlier results on the temperature variation of strains induced in thin silver films at low temperatures. Some problems of differential thermal expansion, and of absolute calibration, remain to be resolved.

Much of the recent work has concentrated on further investigations of the low temperature anomalies in the strain sensitivity of the conductance of silver found earlier when the strain is applied via a surface charge. We have now observed that when the surface is saturated by a liquid, especially such as at the condensation temperature of nitrogen, the measured signal proportional to bulk strain is greatly reduced. This effect adds an additional parameter to the experimental program that has to be controlled, or can be explored further in its own right. For the time being, we will follow the former course, since the basic deviation of the observed signals from what is expected from bulk theory still persists, and requires an explanation.

4. STATE OF THE ART AND PROGRESS DETAILS

Much progress has been made in recent years in exploring the structure of a great variety of materials. The findings confirm that surfaces rarely consist of an abrupt termination of the bulk crystal, as if all matter on one side of a mathematical plane cutting the crystal is removed. Instead, atoms in the surface region respond to the absence of nearest or more distant neighbors on the other side of that plane by rearrangements of position and charge distribution. This leads to adjustments ranging all the way from somewhat modified lattice spacings relative to the bulk crystal to major reconstructions that must be considered as new, quasi-two-dimensional solid phases on top of the surface.

Most experimental probes responsible for such description of real surfaces rely on the underlying assumption that, whatever the rearrangement, surfaces remain homogeneous and periodic. This allows converting the usual primary scattering data into information about the surface periodicity as well as the structural arrangements within each periodic surface cell. Deviations from perfect periodicity generally manifest themselves in line broadening and other smearing out of the data, which, if too large, may in fact make the probe ineffectual. In addition, long range deviations from periodicity are often obscured by the restricted resolution and coherence area of the probing beam itself. Such techniques are designed to look for order rather than disorder.

Nevertheless, simple arguments concerning lattice equilibrium lead to the proposition that many of even the most perfect surfaces should show disturbances in their periodic arrangements that are reminiscent of the existence of faults, such as dislocations, in bulk materials. For example, one often finds increases in lattice spacing relative to bulk of the order of one percent between the outermost atomic layers of a surface region. If interpreted as a uniform strain of order 10^{-2}, this increase should be accompanied by a lateral contraction of the same magnitude (Poisson's ratio), which would then result in a lattice mismatch of one unit cell for every few hundred unit cell displacements along the surface. Actually, though, a strain of order 10^{-2} is far beyond the elastic limit of most materials, and we would therefore expect that, just as in the beyond-the-elastic-limit-response in bulk, the equilibrium structure of the surface will tend to concentrate the strain in small regions of faults, leaving most of the surface strain-free.

The detailed nature of such faults is not clear. They could be vacancies, or line faults, they could cause localized buckling or introduce steps in the surface. For example, surface dislocations were first proposed about 30 years ago,[1] but until recently the level of understanding of most surfaces did not warrant the inclusion of such detail in describing their structure. To the best of our knowledge, there still exist no microscopic or elastic models of the detailed rearrangements for optimum stress relief. But recently, for instance, LEED data on the (111) surface of silver have identified a coherence length associated with the surface of about 75Å,[2] though without further characterizing the details of the structural fault responsible for it. Similarly, the nature of the so-called active sites on a metal surface as possible contributors to various anomalous optical responses of individual molecules attached to a metal surface is receiving increasing attention.[3] Obviously, such faults must also have special importance in the (initial) interaction of clean surfaces with the exterior, ranging from chemical activity, such as corrosion or catalysis, to further crystal growth on the same surface.

For such reasons, basic studies focusing on the structural faults of surfaces rather than on their locally periodic arrangements are timely. Since relatively little is known about this field it makes sense to first concentrate on classes of surfaces that are rather simple, that can easily be brought to thermodynamic equilibrium and are reproducible, that are accessible to probes sensitive to surface faults, and for which an adequate underlying theory exists for their periodic arrangements such that its extension to non-periodic regions can be attempted.

We carry out such a study here, formulated with the above criteria in mind. The surfaces will be single crystal surfaces of the noble metals Ag, Au, and Cu, and the probe will be the conduction electron scattering by the surface, as manifested by the surface-sensitive contribution to the electrical conductance of thin films of these metals. The noble metals were chosen because, by a fortunate coincidence, they can be made to form reproducibly extremely flat and well-ordered surfaces by properly controlled epitaxy on very thin ruby muscovite mica, and because much recent information is available about the theoretically expected charge redistribution and lattice distortion in the surface regions of these metals.[4]

Surface conductance in thin films becomes an obvious probe, because, first of all and in contrast to most other surface probes, it is most sensitive to the <u>disorder</u>, rather than the order of the surface. Secondly, in very thin films surface scattering can be made to dominate all other scattering processes contributing to the film resistance, especially at low temperatures. Since modern measuring techniques for conductance have a sensitivity of a part per million, extremely small changes in surface scattering can be detected. For example we will describe below that a single monolayer of disordered material added to an ordered surface produces changes in resistance at liquid nitrogen temperature of more than 10 percent. Extrapolating, one should be able to see as little as 10^{-4} or less of a monolayer of added material, and a recent study has, in fact, verified this possibility.[5] In addition, modulation techniques, such as by applied surface charge, can be used to provide additional information and to extract further detail at the 10^{-6} relative signal level. The high sensitivity of this probe, if thought of as a scattering technique, obviously arises from the enormous intensity of the incident beam of charge-compensated and effectively non-interacting electrons. Its chief drawback is that the incident and scattered beam directions are ill-defined, and only their averages can be obtained. However, for purposes of this investigation, which concentrates on surface disorder, the averaged cross-sections obtained in this manner will still contain much information. It is, of course, equally important to possess suitable theoretical models of surface scattering for processing the information provided by experiment. As described below, one such model has recently been developed by us as part of the current research unit, and its initial applications have yielded new and fruitful insights into surface structure.

The unique combination of having available high quality surfaces, in a configuration allowing the use of a very sensitive measuring technique, and backed by a promising and relatively straightforward theoretical model, forms the basis of this proposal.

While the use of resistance to monitor surface reactivity has a long history,[6] recent work along lines similar to those we propose is being reported by groups at CNRS-Orsay, France,[5] and at Kharkov University, USSR.[7] However, both programs are only using resistance measurements, while we expect that the simultaneous information provided by surface charge modulation is important for a more complete surface characterization. The direction of recent surface modulation studies at Clemson University[8] is different from ours, and aims at understanding this effect in semimetals where complications of band structure, anisotropy, and small Fermi level have to be sorted out. An important additional difference between our proposed work and that at Orsay is that we are concerned about absolute information on surface structure as well as any changes in it introduced by small amounts of added-on material. Furthermore, our interpretation makes full use of the fact that at low temperatures and in good films, many successive scatterings off the surface under investigation, as well as the opposite film surface, must be taken into account, rather than using the asymptotic formulas of size effect theory that are valid only in the limit where the film thickness is much larger than the mean free path.

In order to understand the results of our most recent work, it is necessary to place them in perspective by including some immediately preceding stages of progress. The most recent results are included below in their appropriate contexts.

The overall aim of the past and current program was to gain a better understanding of the origin of the scattering of conduction electrons at metallic surfaces. After initial work[9] had demonstrated that the combined information of both the electrical conductance and its variation with surface charging, measured over an extended temperature range, could be satisfactorily interpreted in terms of a surface specularity and its modulation with surface charge, the experimental work split into two separate but interacting directions, both aiming at elucidating the origin of this surface specularity. One direction concentrated on clarifying the various new effects discovered during the study of the charge modulation of the surface scattering of well-defined interfaces, and the other focused on methods of controllably producing surfaces of different specularity.

With respect to the charge modulation of metallic thin film conductance, we were able to show for the first time that the two major responses, one proportional to the first power and the other to the second power of the surface charge density, were primarily associated with surface and bulk properties, respectively, of the thin film sample.[10] Further, we demonstrated, also for the first time and by two independent methods, that the quadratic term was related to elastic strains set up throughout the thickness of the sample by electrostatic charging. In one method, we observed the change in the direction of magnetization (of the order of seconds of arc) with surface charging in permalloy films, caused by magnetoelastic coupling.[11] In the other, we were able to suppress the quadratic signal altogether by efficient mechanical clamping of the sample.[10] This quadratic effect, though not directly concerned with surface properties, nevertheless had an important bearing on understanding the linear, surface-related term. The similarity of the two magnetization-dependent responses in permalloy, one linear and one quadratic in surface charge density, practically forced us to interpret the surface-related signal as also coming from strain-mediated changes in direction of the magnetization originating in the immediate surface region.[12] This discovery set the stage for a serious probing of the proposition that most, if not all, surface scattering of conduction electrons is related to localized strain fields emanating from the surface and extending inward over a small distance. Our most recent progress is based entirely on this interpretation.

However, in order to implement an experimental program aiming in this direction, it had to be reoriented towards studying the effects of charge modulation on free metal surfaces rather than on a metal-insulator interface. In both configurations the nature of the observed effects has been found to be essentially the same, but the free metal surface has the advantages of being simpler, and also of being more accessible, even though the overall experiment itself becomes considerably more complex. In addition, the temperature range of our measurements was extended downward to less than 2°K in order to enhance the effect of the surface on transport properties as much as possible. Finally, to support the interpretation of these measurements, a new theoretical model of conduction-electron surface scattering was developed, based on scattering in depth within a thin surface layer that has electrical properties differing from those of the underlying material of the metallic film.[13] The two parameters characterizing this layer are its thickness, and its mean free path.

The most important findings, interpreted in terms of and fully compatible with our theoretical model, are:[14]

SECTION II: SOLID STATE ELECTRONICS

1. The mean penetration depth of surface strains on our well-annealed metal surfaces is about 10 to 15Å, as determined by the thickness of a poorly conducting overlay, such as graphite, that suppresses the linear term in charge-induced scattering (without at all affecting the quadratic term). The decrease of signal with graphite thickness is shown in Figure 1.

2. The surface mean free path of our samples is temperature-independent, and of the order of several hundred angstroms. This is determined by analyzing the (very pronounced) temperature dependence of both the conductance and its modulation with surface charge in terms of our theory. The agreement with charge-modulated scattering and the theory is shown in Figure 2.

3. Detailed application of the theory interprets the surface mean free path as a correlation length for surface order, and thus it also specifies a mean distance between localized fault regions.

4. Both the linear and quadratic responses show anomalous behavior at temperatures below 12°K, with the signals tending to disappear. This is most pronounced for the quadratic term which drops precipitously, sometimes below the noise level, below 7°K, as shown in Figure 3. At this writing, we do not know whether this effect arises from a suppression of strain-sensitive scattering, or from a shielding of charge-induced strain from the bulk of the sample.[15] Either interpretation would have to invoke new and so far never-discussed properties of metals or charged metal surfaces at low temperatures.

According to the first finding, the distance below a metallic surface within which electric fields applied to the surface can affect electrical conduction is apparently much larger (~ 15Å) than the distance of the electrostatic screening (~ 0.5 Å) of the electric field itself. This conclusion was reported last year, and was based on studies of the influence of a thin layer of graphite above the metal on the surface scattering of current carriers in the presence of an electric field applied to the surface of a thin film of silver. The typical decay of this influence as a function of graphite thickness is shown again in Fig. 1. Carbon has the property of covering the silver extremely uniformly, and it undergoes little diffusion into the metal. In parallel with the silver, it does not contribute measurably to electrical conduction, nor does it induce scattering at the silver-carbon interface, as deduced from the fact that there is no change of electrical resistance upon carbon deposition. Its conduction electron density produces a screening depth of about 2 Å to 5 Å, considerably less than the 15 Å observed in Fig. 1, which must therefore reflect non-electrostatic, and presumably elastic, effects.

In order to further understand this phenomenon, we have extended these studies to other layer structures, with screening lengths larger than that of graphite. First of all, we have confirmed that the results of Fig. 1 apply whether the carbon is between the silver and a dielectric, or when it is on the free surface of silver. This latter possibility is

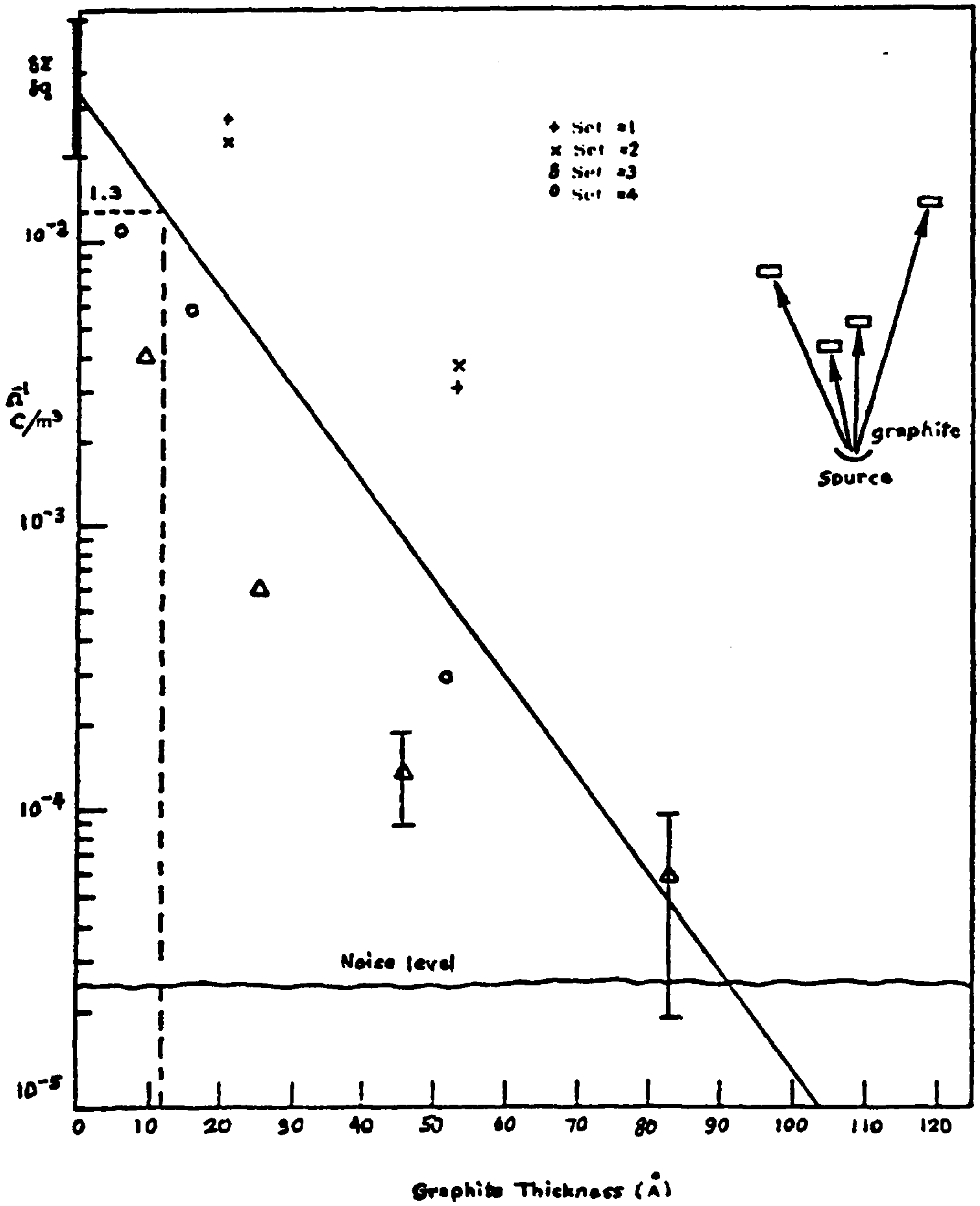


Fig. 1 Variation of surface charge density modulation of silver film conductance, $\delta\Sigma/\delta q$, with thickness of graphite overlay, at a silver-mica interface.

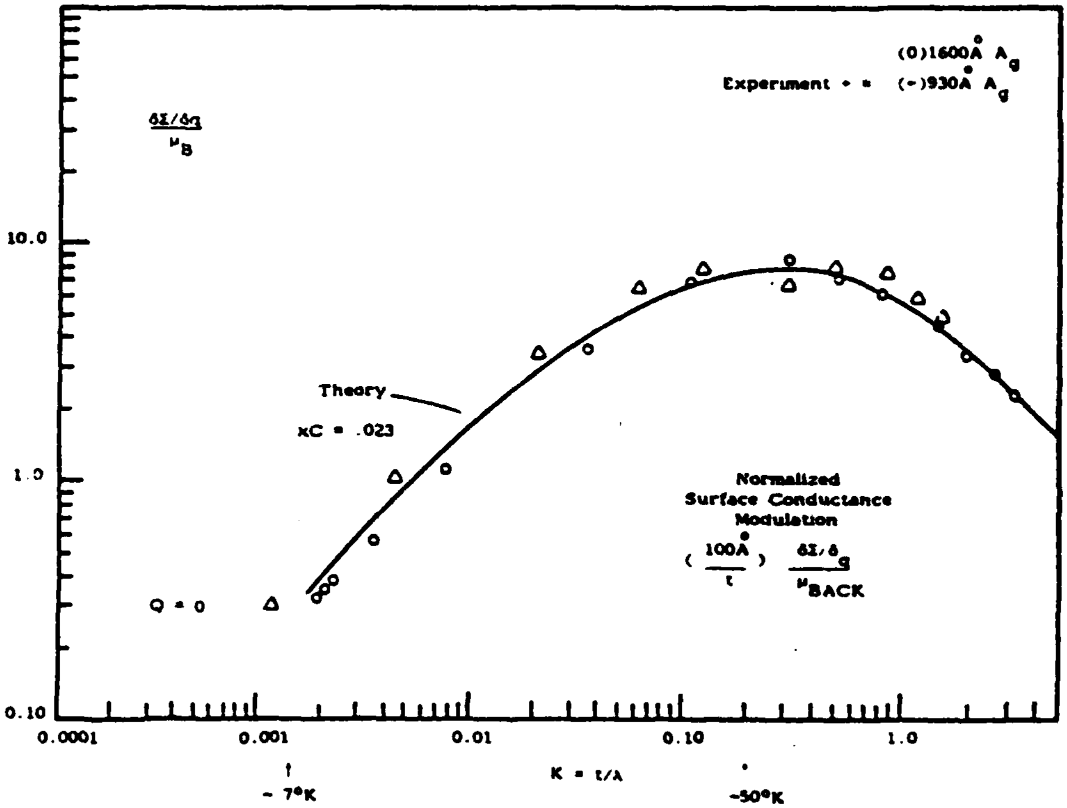


Fig. 2 Temperature dependence of $\delta\Sigma/\delta q$ at the free metal surface of two silver films, relative to the bulk mobility μ_B, and reduced to a 1000Å thickness. The horizontal scale, in terms of the universal size effect parameter K (= thickness/bulk mean free path), is directly proportional to the absolute temperature T. The experimental points follow the line based on the theory of surface scattering caused by surface charge-induced modulation of surface disorder.

important because this geometry allows a direct examination of the carbon overlay, such as by an Auger probe. We have detected only trace amounts of silver above or through the thin carbon layer, and certainly much less than the amount needed to reproduce the data of Fig. 1 on the basis of an inhomogeneous coverage alone. The penetration depth of ~ 15 Å persists, and the insensitivity of the Maxwell str‑ ‑ses to the presence of the surface layer, reported earlier for a sil -dielectric interface separated by a graphite layer, continues to e . at the free surface.

The extension of this approach to other possible metallic layers of intermediate screening depths has not been too successful. Obvious candidates, such as antimony, give complicated results because they interdiffuse readily, and because they also produce noticeable interface scattering. Attempts to separate the current carrying silver film from a thin metallic overlay by an insulating barrier, such as Al_2O_3, show promise, but may have to be produced and measured entirely at low temperatures in order to minimize diffusion through the barrier in either direction. In addition, it still has to be established how uniform a metallic overlay can be produced on top of the alumina barrier. It may very well be that graphite is unique in satisfying all the requirements of this particular experiment. In that case, it becomes important to devise a corroborative experiment for the screening length of our graphite layer before the interpretation of Fig. 1 can be firm.

In consonance with the work outline of the proposal, and in anticipation of the need to extend further experiments to very low temperatures, we are constructing a new experiment for measuring the strain sensitivity of the resistivity of bulk metals at low temperatures. Information on this behavior does not exist in the literature, and since this behavior is needed to interpret the conductance modulation data due to Maxwell stresses at low temperatures, we have undertaken to design an experiment to determine the needed experimental data.

We have now established that there is sufficient sensitivity to measure this effect reliably with the same apparatus used to detect the surface conductance modulation of silver films. The major remaining design problem before embarking on the measurements themselves concerns the differential thermal contraction of various components of our sample holder and sample. Since one of the objects of carrying out these measurements is to establish the onset of higher order terms in the strain sensitivity of resistance at low temperatures, we must at all times know the operating point around which the sample is being strained. As this changes with temperature if there is a substantial differential thermal contraction, we are devising another set of piezoelectric controllers to give information on, and, if necessary, to independently vary, this operating position.

The calibration of all piezoelectric transducers over an extended temperature range also remains to be established in order to put the observed effects on an absolute basis.

During the past year, we have reinvestigated the effects of strain on the conductance of silver films at low temperatures, such as shown in Fig. 3, with two aims in mind. First of all, to see, by varying the

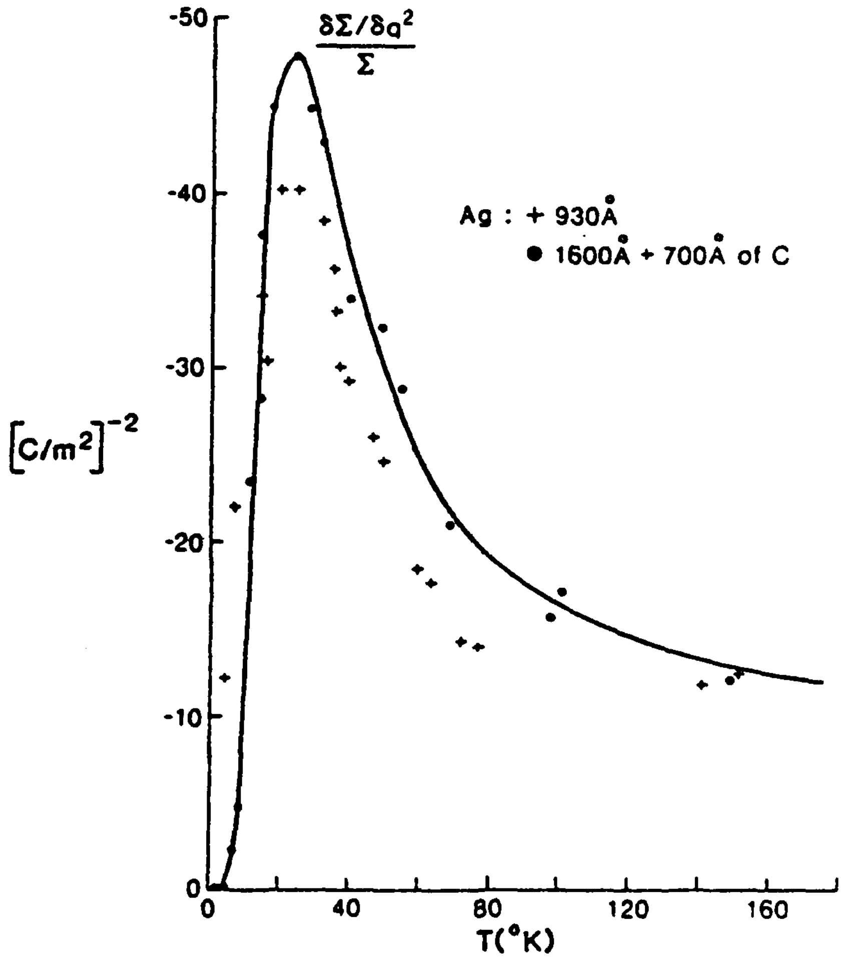


Fig. 3 Relative film conductance modulation proportional to q^2, as a function of temperature for two samples: a 930Å Ag film charged on its free surface (+), and a 1600Å Ag film charged on the outside of a 700Å graphite overlay (·). Note the anomalous peak followed by a steep drop at very low temperatures.

specific experimental procedure, that these results are intrinsic in the material and do not depend on details of experimental approach, and secondly, to confirm the overall behavior on samples prepared independently and under deliberately slightly different conditions.

The results of this work, concentrating on the temperature range T>60°K, are that, indeed, the upward trend of the strain sensitivity observed in Fig. 3 is fully reproducible in this temperature range, but only if two conditions are carefully maintained. Namely, care must be taken to eliminate any possibility of mechanical clamping of the sample, such as, for example, as a result of differential thermal expansion of parts of the sample holder, or of the sample itself. In addition, the signal is appreciably reduced relative to the points on Fig. 3, in the same temperature range, when the sample surface is exposed to condensing liquids. This is particularly noticeable when the sample is in a nitrogen atmosphere of the liquid nitrogen coolant. The same effect can also be induced by other liquids condensing at other temperatures. Since the original experiments of Fig. 3 were carried out in an environment above a liquid helium bath, it now appears very likely that the disappearance of the signal near $T \simeq 4°K$ in Fig. 3 can be caused by the same effect. This is not easy to prove because of the difficulty of keeping He away from the sample in this temperature range. Why this occurs is not clear, at present. Possibilities range from a decoupling of the electric field effects from the interior because of the liquid layer, or because of diffusion below the surface, to inertial loading of the sample by the liquid that prevents its deformation. In any case, clarification of this point has now put the data for T>60°K on a firm footing, and we will now proceed on further examining the causes of the rise of the signal in Fig. 3 above its room temperature value, in this region of temperatures.

It is significant to note that throughout this period of experimentation, the signal that we have been ascribing to surface strain distortion by charging, as shown in Fig. 2, has remained completely unaffected by the presence of any condensing liquids. In this connection, it is also worth mentioning that our conclusion on the basis of these experiments (Figs. 1 to 3) that even a perfect surface, such as Ag 111, has sufficient strain in it to cause long range deformations, such as buckling, has now been confirmed by electron micrograph examination of these surfaces.[16] The characteristic length, in good agreement with our results, is of the order of a few hundred Å.

Both the work related to Figs. 1 to 3, and the work on artificially roughened surfaces that is now complete, are being prepared for publication.

5. REFERENCES

1. C. Herring, "Surface tension as motivation for sintering," in The Physics of Powder Metallurgy, McGraw-Hill, New York (1951).

2. D.G. Welkie, M.G. Lagally and R.L. Palmer, "LEED study of the surface defect structure of Ag(111) epitaxially grown on mica," J. Vac. Sci. Technol. 17, 453 (1980).

3. A. Otto, I. Pockrand, J. Billman and C. Pettenkofer, "The ADATOM Model: How important is atomic scale roughness?" in Surface Enhanced Raman Scattering, eds. R.K. Chang and T.E. Furtag, Plenum (1981).

4. U. Landman, R.N. Hill and M. Mostoller, "Lattice relaxation on metal surfaces, an electrostatic model," Phys. Rev. B21, 448 (1980).

5. J.P. Chauvineau, "Diffusion superficielle des adatomes observée a basse temperature par la variation de résistance électrique de films minces d'or," Surf. Sci. 93, 471 (1980).

6. Chemisorption and Reactions on Metallic Films, J.R. Anderson, ed., Academic Press (1971).

7. I.P. Grebennik, "Resistivity of polycrystalline silver film with artificial surface roughness," Phys. Met. Metall. 47, 42 (1980).

8. D.A. Glocker and M. Skove, "Field effect and magnetoresistance in small Bismuth wires," Phys. Rev. B15, 608 (1977).

9. A. Berman and H.J. Juretschke, "Size effects in resistivity of epitaxial films of silver," Phys. Rev. B11, 2893 (1975).

   A. Berman and H.J. Juretschke, "Origin of the metallic field effect," Phys. Rev. B11, 2903 (1975).

10. D.J. Lischner and H.J. Juretschke, "Maxwell stresses at charged metal surfaces from thin-film elastoresistance," J. Appl. Phys. 51, 474 (1980).

11. P. Mazumdar and H.J. Juretschke, "Magnetoconductivity modulation of thin ferromagnetic films by substrate electrostriction," Phys. Rev. B19, 664 (1979).

12. P. Mazumdar and H.J. Juretschke, "Oscillatory metallic field effect and surface magnetoelasticity in thin ferromagnetic films," Phys. Rev. B19, 672 (1979).

13. H.J. Juretschke and R. Pimpinella, "Size effect in thin metallic films resulting from modified surface transport properties," submitted to Surface Science.

14. R. Pimpinella, "Metallic field effect, surface scattering and surface structure of free metal surfaces," Ph.D. Thesis, Polytechnic Institute of New York (Fall 1981).

15. J.M. Lockhart, F.C. Witteborn and W.M. Fairbank, "Evidence for a temperature-dependent surface shielding effect in Cu," Phys. Rev. Lett. 38, 1220 (1977).

16. L.D. Marks, V. Heine and D.J. Smith, Phys. Rev. Lett. 52, 656 (1984).

C. X-RAY COUPLED WAVE INTERACTIONS AT CRYSTAL SURFACES

Professor B. Post and H.J. Juretschke

Unit SS4-3

1. OBJECTIVE(S)

To use the multiple interaction of x-rays in a Bragg geometry in order to develop simple and compact methods for obtaining direct phase information about the crystal scattering factors, and for characterizing the mode structure; to extend the interactions to include coupling to other waves that can modify the mode structure, and that can be used to explore nonlinear interactions of x-rays; and to understand the local x-ray fields in the immediate surface region of the solid, as well as the effect of stringent boundary conditions on all modified x-ray waves, especially those originating in the interior.

2. APPROACH

High resolution x-ray diffraction in the region of n-beam interactions contains both phase and mode structure information and will also reveal effects of coupling to other waves by specific angular shifts in the diffraction directions. The experiments will be supported by theoretical studies of the predictions of n-beam dynamical theory, and by rigorous extensions of the theory to include the coupling with other waves.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of the most recent progress; earlier annual reports and the original proposal provide background and prior status of the various problems, and more detailed descriptions are contained in the next section in conjunction with the state of the art, so that the nature of the contributions can be understood more clearly.

A. n-Beam Interactions

The analytic first order theory of n-beam effects on a two-beam Bragg reflection has been refined further both in order to gain additional physical insight into the origin of the initial modifications as the n-beam is approached, and also to set the stage for the development of the next order of the theory. To measure the validity of our first order solution, it has also been compared to exact solutions that bear out the conclusions we have reached on the basis of our formulation.

B. X-ray Interaction With Thermal Phonons

An earlier discrepancy between the theory of thermal diffuse scattering developed by us and available experimental data has now been resolved by further analysis of the experiment that shows the observed signal to be incompatible with the magnitude of x-ray phonon interactions. At the same time, our dynamical theory is in good agreement with recent observations of defect lines crossing x-ray spots of diffuse scattering, but these observations have not yet received a detailed theoretical interpretation.

C. Direct Phase Determination

The practical rules established for directly determining phases of invariant combinations of structure factors in Renninger-type experiments have been tested extensively on a variety of interactions and crystals, and have shown full agreement with the values traditionally obtained by indirect analysis. This work has received mention in the annual review of physics by the American Institute of Physics.

4. STATE OF THE ART AND PROGRESS DETAILS

With the advent of new high intensity sources covering a continuous frequency spectrum, such as the synchrotron, x-rays are assuming a significant role as a tool in surface exploration, similar to their traditional role in the structure determination of bulk solids. At the same time, the new degrees of freedom opened up by these sources now allow including x-rays as a fully accessible part of the conventional electromagnetic spectrum, with the same possibilities for application that have been obtained historically, for example, in the microwave and visible parts of the spectrum. From this point of view, however, both the conventional treatment of x-ray interactions, as well as the experimental techniques being used, have to be reexamined and redirected in order to make them suitable for exploiting these new directions.

The basic interaction of x-rays with crystals, as formulated more than fifty years ago, has always included the possibility that more than the two x-ray fields of common diffraction phenomena, the so-called incident and diffracted beams, can couple within the crystal.[1] Such more general coupling, designated as the n-beam case, however, occurs only under stringent conditions of geometry and wavelength, and is observable primarily in crystals of good perfection and when using x-ray sources of high monochromaticity and high angular definition. Among the most startling consequences of these n-beam interactions have been (a) the phenomenal decrease in linear absorption of beams traveling along some special directions, allowing highly collimated x-rays to pass literally through centimeters of normally highly absorbing material,[2] and (b) the fact that n-beam interactions can be used to extract the phases of individual scattering factors,[3,4] information that has been one of the long sought goals of crystallographers because it is needed to put the determination of molecular structure from measured diffraction intensities on a fully deterministic footing. The underlying theory of these effects, the dynamical theory, is fully formulated, and various aspects of the properties of some special 3- to 6-beam interactions have been investigated both experimentally and theoretically.[5-7] But the complexity of the theory when applied to n strongly coupled beams and the increasing diversity of possible experimental arrangements for observing the effects have, so far, prevented the design of a truly optimal and efficient experimental technique that is applicable to a wide range of crystals and n-beam combinations. At the same time, there is still no formulation of a straightforward and unambiguous method for interpreting n-beam data with respect either to the propagation characteristics of particular x-ray beams, or to phase information about a structure.

The same stringent experimental conditions needed for the appearance and study of n-beam interactions also allow detailed observation of the coupling of x-rays to other types of waves propagating within the crystal. These waves, such as for instance, phonons, photons in the visible, and polaritons, usually have energies small compared to x-ray photons. But they can contribute enough momentum to the interaction to produce small but observable angular shifts in and out of a particular n-beam interaction region. In principle, then, such other waves can steer x-rays, and can therefore perhaps control their propagation characteristics, or their absorption within the crystal host. They are, of course, also of interest in their own right in elucidating higher order couplings and nonlinear interactions in crystals. The experimental work in this area has usually been confined to first order interpretation. Most of the theoretical work has concentrated on applying the conventional methods for treating nonlinear interactions in matter without taking into account the specific nature of the x-ray fields of interest.[9] This includes the fact that x-rays are dispersive in the configurations of most interest when interacting with other types of waves. For example, one usually ignores that x-ray fields form coupled plane wave modes even when the conditions for diffraction are not fully satisfied, so that the initial and final states of the x-ray in a scattering process are not plane waves, but rather Bloch waves, composed of two or more coherent plane waves of fixed relative amplitudes and propagation vectors. Obviously, since the effects of such scattering are only observable outside the crystal, the outside plane wave modes must also relate properly to the inside coupled modes through rather intricate boundary conditions before reliable predictions of observable phenomena can be made.

Since both the interactions of x-rays with each other, and the interactions of x-rays with other types of waves, call for essentially the same theoretical formalism, and for similar conditions for experimental observation, they together form a natural combination in the research project that we propose here. The work will concentrate on such interactions in the neighborhood of the surface of entry of the incident x-ray beam, to be observed by x-rays emerging from the same surface. This is the so-called Bragg case, and it has been chosen partly because the experimental observations are more sensitive for reflected beams than for beams transmitted through a crystal slab (the Laue case), and partly because the mode boundary conditions only involve one interface, for a sufficiently thick crystal. A careful experimental investigation under such an arrangement requires, of course, not only that the crystal structure be very good on the whole, but also that the surface region of interest have comparable regularity, so that homogeneity persists throughout most of the region of interaction.

The work that we pursue has two major specific goals: (a) to develop simple and straightforward procedures for determining the phases of scattering factors in complex crystals and of less than perfect order; (b) to establish both experimental and sound theoretical criteria for the systematic study of the interaction of x-rays with various other kinds of waves. As we will see below, these two goals will, in fact, involve many other aspects of coupled wave interactions, including those already mentioned above.

In order to understand the results of our most recent work, it is necessary to place them in perspective by including some immediately preceding stages of progress.

Since the concerns in the first two main areas, n-beam interactions and coupling with other waves, involve different levels of approach -- the first immediately drawing on a full-fledged theory while the second begins with very little sound theoretical foundation -- the two areas will be discussed in sequence. It will soon become clear, though, that both are intimately interconnected.

A. n-Beam Interactions

Largely for historical reasons, because the first indication of anomalous interaction features (Borrmann effect) had been found in the propagation through crystals, our work concentrated on experimental consequences in the Laue case, which detects the diffracted beams moving in the forward direction relative to the incident beam, beyond the exit surface of the crystal. This configuration was also employed in the seminal paper on phase determination in 3-beam interactions by Professor Post[3] that formed the basis of much of our subsequent work. A number of specific 3-, 4-, and 6-beam Laue interactions have been studied in detail,[5,6] partly in order to understand the development of their mode structure as one passes through the n-beam region along specific directions, and partly in order to try to abstract those common characteristics that should allow for at least a qualitative prediction as to what happens in other cases. Most of the studies employed computer experimentation because some of the constraints for observing effects of coupling were not known initially -- and subsequently turned out to be beyond those of x-ray sources other than the synchrotron, so that actual experimentation would have been inconclusive. The major results which emerged from these studies were:

(1) The mode structure is an extremely sensitive function of location within the n-beam region, with relative amplitudes, propagation vectors (complex), and polarizations of the various plane wave components varying very rapidly with changes of seconds of arc.[6,10]

(2) For any specific n-beam case, the polarization of the incident beam is an important variable for optimizing the amplitudes of the least absorbing components of the n-beam mode, and therefore also optimizing the transmitted intensity.[6]

(3) The n-beam interaction can be used to produce highly collimated pencils of x-rays (divergence of a second of arc or less) with specified polarization. This suggests it as a highly selective optical component for x-ray experiments at synchrotron facilities.

(4) In contrast to the complications of the mode structure, the dispersion surface in the n-beam region shows many regular features, mostly dictated by geometry and continuity, which allow both an immediate qualitative identification of the predominant mode components, as well as an identification of the invariant phase factors associated with the particular interaction.

(5) Applying the ideas of (4), phase information can be obtained by examining the asymmetry in the intensity of the transmitted beams as one passes through the n-beam region.[11]

(6) While the procedure of (5) has been successful for nearly perfect crystals, such as Ge and Si, our experimental work on less perfect crystals, especially organic ones, has not given unambiguous results. From this work, however, it seems plausible that if the same investigation is carried out at longer x-ray wavelengths such as supplied by synchrotrons, phase effects should become more pronounced.

These results have now been extended systematically to the Bragg case, again looking for the left-right asymmetry in diffracted asymmetries as one passes through n-beam regions along a weakly reflecting 2-beam line. The data show pronounced and systematic asymmetries under only moderately stringent experimental conditions, for practically all the n-beam interactions crossed by the two beam line, with the asymmetry in the simple cases always in the direction predicted by inspection of the dispersion surface. Because of its ease, this method is a practical one that would make phase determination routinely possible.

Our results this past year on n-beam interactions are the following.

The theory of n-beam interactions, formulated analytically as a perturbation expansion around a 2-beam solution, and first reported last year, has been refined further. A long paper, summarizing our findings, is in press.[26] Apart from confirming theoretically that information about phases of invariant combinations of structure factors can be obtained from the wings of the lines in Renninger-type experiments, the analytical theory also develops other general properties of n-beam solutions close to the n-beam point that have so far not been appreciated from the findings of specific computer-generated solutions. In particular, it has identified a region near, but not at, the n-beam point where the proper modes of x-ray propagation are rapidly changing as a function of angle of propagation of the "incident"beam, and their description may be more complicated than that at the n-beam point itself. This region corresponds to the x-ray analogue of band crossings for electrons in solids. As an example, we show in Fig. 1 the edges, in reciprocal space, of the regions of total reflection for the Ge $222/11\bar{1}$ interaction as a function of the reduced azimuthal angle x, resulting from first order analytical theory. Crossings occur at $x=-1$ for the solid lines (σ-polarization), and at $x=-1.5$ for the dashed lines (π-polarization). At these crossovers, the effective structure factors vanish, thus giving minima in x-ray reflections. That this first order theory is quite adequate for the description of the physics in this region can be seen from Fig. 2, which presents an exact solution of the same interaction. While one of the crossovers has been removed, all the structure factors, measured by the spacing between lines belonging to the same mode, still show pronounced minima in this region. But the edges of the totally reflecting regions are very complicated, and, in fact, show discontinuous changes in the region close to $x=-1$. In order to see these effects, of course, one has to be able to excite just one of the normal modes of propagation, which is difficult, since, obviously, the polarization of the proper modes undergoes very rapid changes in the same region. The identification of this region may allow breaking down the general n-beam problem into two subproblems, one near $x=-1$, the other near $x=0$, as a way of developing a general solution. Incidentally, our first order solution can be shown to be a generalization of a perturbation approach developed for describing electron diffraction in terms of "dynamic potentials."[27] The generalization consists of treating transverse rather

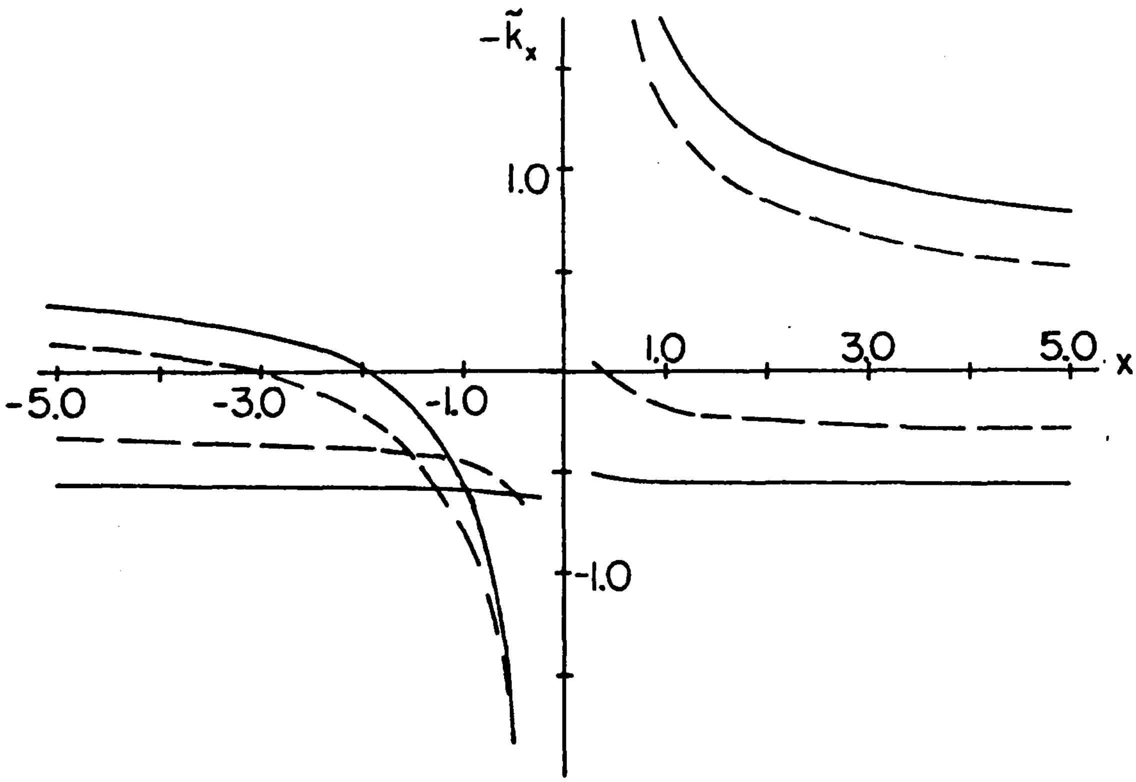


Fig. 1 First order analytical approximation of the edges of the regions of total reflection in reciprocal space of Ge 222/11$\bar{1}$ as a function of reduced azimuthal angle x. The solid and dashed lines refer to the two independent polarization states that asymptotically become σ- and π-plane polarized.[26]

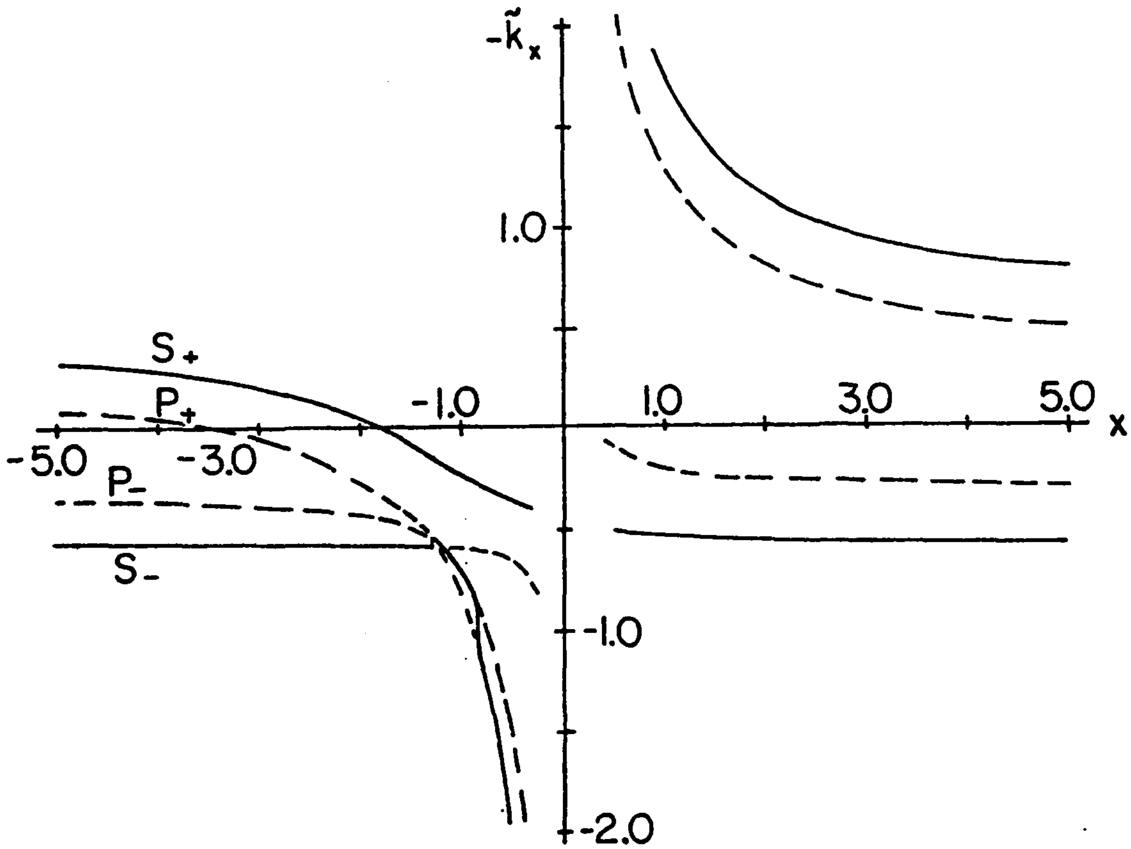


Fig. 2 Exact solution of the interactions in Fig. 1.

than scalar waves, with its complicating factors of doubling the mode structure and interpolarization couplings, and of applying it to the specific configurations of importance in x-rays, as well as of a more complete specification of the range of validity of this solution.

B. Interactions of X-rays With Other Waves

As already mentioned with respect to this topic, one of the major thrusts during the current program was directed towards extending the conventional dynamical theory of x-rays in crystals to systematically include the coupling with other waves within this framework. We immediately chose the Bragg case for x-rays, for reasons like those already mentioned, and confined ourselves entirely to the two-beam case in order to learn how to handle the desired coupling in the least non-trivial dynamical setting. As with all such theoretical probing tied to a simple physical model, we developed a number of physically motivated approaches to the problem before formulating a general theory. In order to give the theory concrete content, we focused on furnishing a quantitative interpretation of two experiments in the recent literature. In one experiment, satellite diffractions were observed when intense beams of phonons travel parallel to the reflecting surface.[12] In the other, details of the fine structure of x-ray-thermal phonon scattering were displayed in an arrangement in which the incident beam entered away from the Bragg angle and the coupling produced outputs in a whole range of directions.[8] A third focus of the theory centered on predicting the consequence of the interaction between x-rays and lasers in ionic crystals via the polariton modes, and used this prediction to set up the appropriate experiment. Progress in all of these directions has been very good. However, rather than recount the sequential development through the course of the program, the summary below will list the current state of this investigation. Some of the work is related to a Ph.D. thesis completed in 1982.[13]

(1) We have found a fully analytic solution of the six-beam case of an x-ray interacting dynamically with monochromatic high intensity phonons moving parallel to the crystal surface.[14] We can therefore exactly solve the boundary condition problem, and determine explicitly all external reflected fields for arbitrary intensity of the phonon beam. This exact solution has been extremely useful in exploring the effect of changing the various parameters in the problem, and especially under what conditions the usual weak coupling approximations hold.

(2) We have formulated the dynamic n-phonon problem for all one-phonon processes of n phonons, and set up the explicit equations for the 2n + 1 sheeted dispersion surface and the associated fields. Introduction of the boundary conditions leads to a 2n by 2n determinant for the relative incident fields, and the corresponding expressions for the reflected fields. When the coupling is weak, these expressions reduce to explicit modifications of the uncoupled solutions for all but the main reflection, which has still to be solved for a modified dispersion relation. Our expressions so far show that this modification contains parts equivalent to the Debye-Waller factor but also other contributions that only appear when the full problem is considered. In particular, the theory also predicts a decrease in peak and a narrowing of the main beam, with satellite structure in the wings. Applications to detailed experimental results remain to be carried out.

(3) The interaction of x-rays with laser-induced polaritons has been analyzed from the point of view of energy transfer, and it is found that under optimal conditions, using normal laser intensities, the weak coupling approximation for the phonon part of the polariton is justified. This has made it possible to adapt directly the formalism worked out under (1), above, to single phonons traveling in other directions, but with the additional condition that energy transfer between polariton and x-ray is not completely negligible. This implies that the coupled modes contain fields of differing frequencies that refer to *different* dispersion surfaces. However, our formalism lends itself readily to incorporate this complication. One immediate consequence is that the expected 6-beam coupling reduces to 4-beam cases, i.e., polaritons are predominantly either absorbed or emitted, but not both. We are in the process of determining the optimum experimental configuration for observing laser induced switching of x-rays in a selected material, such as MgO, with the actual experiment being one of the tasks of the proposed research work.

The results obtained this past year related to x-ray interactions with thermal phonons are the following.

The above theory of n-phonon interactions had been applied earlier to thermal diffuse scattering (TDS), i.e., the background intensity in the neighborhood of a Bragg peak caused by x-ray thermal phonon inelastic interactions.[15] That theory led to very specific predictions of the TDS fine structure, of the type shown in Fig. 3, and comparison was attempted, as mentioned in the last report, with the corresponding available experimental data.[8] However, the lack of more than general agreement of the shape of the signals with the theory, which had been ascribed to the effects of a relatively high background noise level in the experiment, gave an unsatisfactory correlation, and led us to withhold publication of our results.

A more detailed analysis by us has now concluded that these experiments, in fact, did not see thermal phonons, but rather most likely a distribution of static imperfections in the surface region of the sample crystal with a Fourier spectrum similar to that of phonons. The signal expected from thermal phonons is more than two orders of magnitude lower than that reported experimentally! One of the additional experimental difficulties is that in order to detect such a small signal due to inelastic scattering, the elastically scattered intensity must be smaller still. This requires, for example, that the wings of the incident beam must be reduced by better than 10^5 within about a minute of arc from the maximum, and apparently even the best crystal monochromators fall far short of this criterion, for reasons that bear investigating in their own right. Nevertheless, it is gratifying that in completely different experiments, where the incident beam is more than a degree away from the Bragg angle, the defect line predicted in Fig. 3 near $\theta=0$ has, in fact, been seen recently[28] and been identified with x-ray phonon interactions, although without detailed theoretical explanation. An extension of our theory to follow the structure of this defect line when the incident beam is within the region of total reflection is in progress.

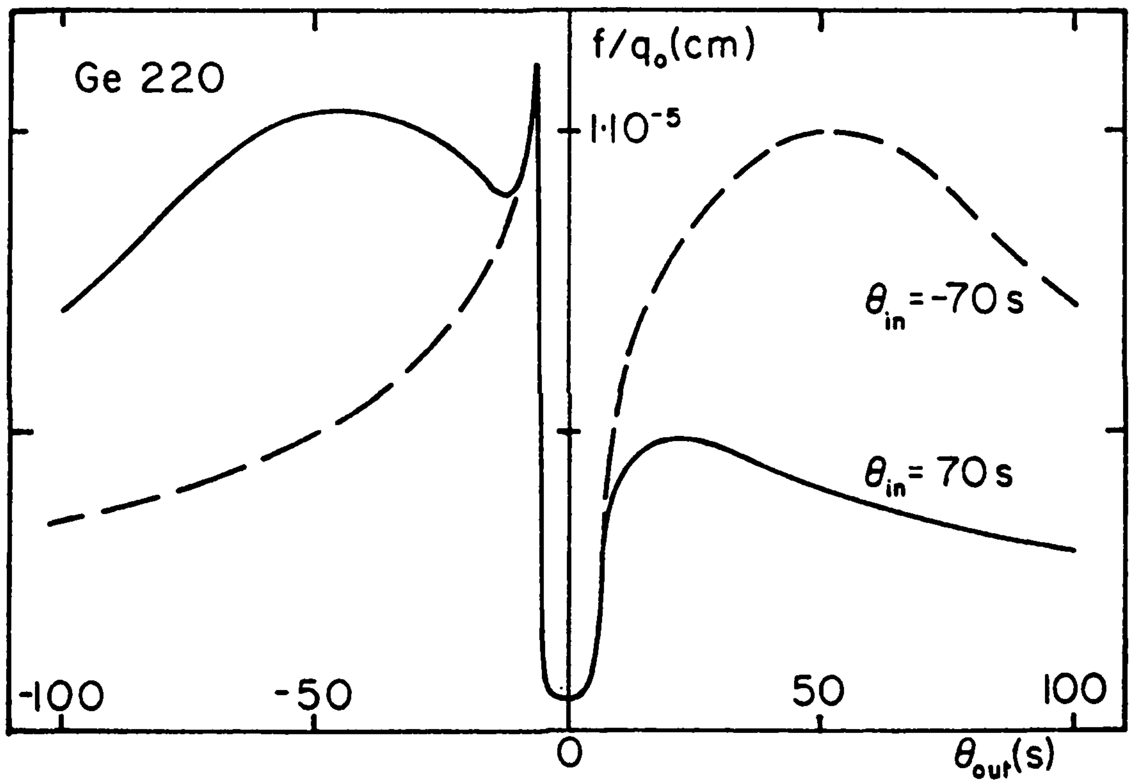


Fig. 3 Thermal-phonon-modified x-ray diffraction for Ge 222. The incident x-ray makes an angle of ±70 seconds of arc with the Bragg angle, and the graphs show the diffracted intensity into different output angles. The reduction in intensity near $\theta_{out}=0$ follows from dynamical selection rules.

C. Phase Determination in Mosaic Crystals

A crucial question about the practicality of observing phase information by the methods we are pursuing is the extent to which the necessary effects persist when dealing with imperfect crystals, because very few crystals can be grown to the requisite purity and perfection so that the full dynamical theory of x-ray diffraction applies. Various corrections have to be made that, in a number of instances, such as the absolute total diffracted intensity, lead to qualitatively different responses in perfect and imperfect (mosaic) crystals.

To obtain answers to this question we have carried out a primarily experimental program that has tested out the appearance of the fine structure in Renninger-type experiments on a number of centrosymmetric crystals of average quality. When the emphasis is focused on minimizing the divergence of the incident beam in all directions, on monochromaticity, and on studying the details around the base line near the n-beam interacton rather than at the full interaction point itself, our work shows that there is sufficiently clear-cut information in the intensity asymmetries to allow a determination of invariant phases in most interactions, at least in centrosymmetric crystals.

As an example, our results on $ZnWO_4$, analyzing 53 interactions, of which 40 represent 4-beam couplings, agree fully with the phases of the known structure. An example[31] of the fine structure obtainable under the appropriate conditions is shown in Fig. 4.

The immediate basis for identifying phases in such crystals has been the correlation between the relative orientation of the maxima and minima in the integrated intensity on both sides of the interaction point, with changes in the absorption coefficient of the most strongly excited sheet of the dispersion surface in the same regions. An explanation of this approach, and its justification in terms of a qualitative interpretation of pertinent features of the dispersion surface in the vicinity of the n-beam interaction point, has now been published.[32]

Most recently, a considerable amount of experience and data has been accumulated to test the limits of applicability of our method for identifying experimentally the phases of various invariant structure factor combinations, in a variety of crystals of differing perfection. Using beam configurations with angular resolutions as low as 4" of arc, the azimuthal angular fine structure in Renninger-type experiments has been explored extensively, and in all cases the phases determined directly experimentally agree with those found by traditional indirect means. We now have enough confirmation in order to apply the method with confidence to unknown phases.

A refined repetition of our results for Ge 222 as the monitoring reflection has been accepted for publication in Acta Crystallographica, and reports on other aspects of this problem have been submitted for publication.[29] A general summary of our techniques has been included in the annual review prepared by the American Institute of Physics, Physics in 1983,[30] and the most recent talks on it are being presented at the International Congress of Crystallography at Hamburg, Germany, in August 1984.

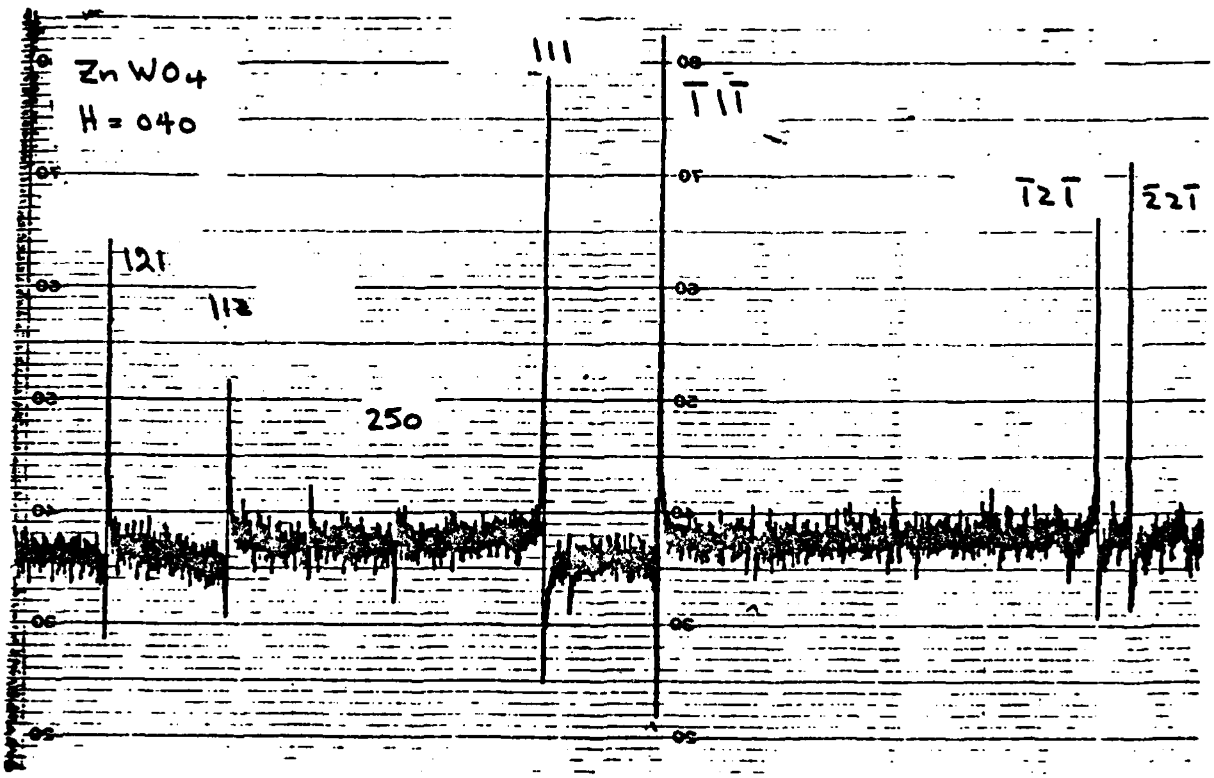


Fig. 4 Experimental chart of the same type as in Fig. 4, but for the 040 reflected intensity of $ZnWO_4$, a crystal of average crystallographic quality (mosaic crystal).

D. X-ray Interactions with Lasers

Among the other results, the analysis of x-ray interaction with lasers via polaritons has been completed, and a computer search is under way for optimizing the observation of this interaction relative to crystal host, to laser wavelength, and to angular shift of the laser- controlled beam. Incidentally, one of the results that has emerged from the study of x-ray phonon interactions, namely the existence of new selection rules suppressing the role of phonons within or close to an x-ray reflection, also affects this interaction, and eliminates the most obvious configuration of x-rays and lasers in producing coupling. The search must therefore concentrate on the wings of the x-ray line, and especially on that side that is characterized by a small absorption coefficient.

5. REFERENCES

1. B. Batterman and H. Cole, Rev. Mod. Phys. 36, 681 (1964)

2. G. Borrmann, Trends in Atomic Physics, ed. O.R. Frisch, Interscience, N.Y. 1959.

3. B. Post, Phys. Rev. Lett. 39, 760 (1977).

4. B. Post, Acta Cryst. A 35, 17 (1979).

5. B. Post, S.L. Chang, T.C. Huang, Acta. Cryst. A 33, 90 (1977).

6. T. Hom, "Three Beam Dynamical Interactions," Ph.D. Thesis, Polytechnic Institute of New York (1979).

7. A. D. Fofanov, A. V. Kuznetsov, V. G. Razgulyaev, Sov. Phys. Cryst. 21 15 (1976).

   A. M. Afanasev and V. G. Kohn, Acta. Cryst. A 32, 308 (1976).

   M. Umeno, Phys. Stat. Sol. (a) 37, 561 (1976).

8. P. Eisenberger, N. G. Alexandropoulos and P. M. Platzman, Phys. Rev. Lett. 28, 1519 (1972).

9. P. Eisenberger and S. L. McCall, Phys. Rev. Lett. 26 684 (1971).

10. P. Wang, "Direct Experimental Detection of X-ray Phases from Intensity Measurements" Ph.D. Thesis, Polytechnic Institute of New York, (1979).

11. P. Gong, "An Experimental Determination of the Phases of Mosaic Centrosymmetric Crystals," Ph.D. Thesis (January 1983).

12. S. D. LaRoux, R. Colella and R. Bray, Phys. Rev. Lett. 35, 230 (1975).

13. F. Robbins, "X-ray - Phonon Interaction as a Dynamical N-Beam Problem," Ph.D. Thesis, Polytechnic Intitute of New York, (June 1982).

14. H.J. Juretschke and F. Wasserstein-Robbins, "The Role of X-ray Boundary Conditions and Other Effects in Strong Dynamical X-ray Phonon Interactions," Phys. Rev. B26, 4262 (1982).

15. F. Robbins and H. J. Juretschke, "Dynamic Effects in X-ray Thermal Phonon Interaction in Symmetric Bragg Reflections," to be submitted to Acta Cryst.

16. P. L. Cowan, J. A. Golovchenko and M. F. Robbins, Phys. Rev. Lett. 44, 1680 (1980).

17. M. Laue, Röntgenstrahl - Interferenzen Akademischer Verlag, Frankfurt (1960).

18. J. Nicolosi, "Experimental Procedures for Determining the Invariant-Triplet Phases of X-ray Reflections," Ph.D. Thesis (June 1982).

19. B. Post, "The Experimental Determination of X-ray Reflection Phases," Warren Award Lecture, ACA Meeting, Gaithersburg, Md. (March 1982).

20. B. Post, "Practical Aspects and Implications of Experimental Phase Determination Procedures," ACA Meeting (see above), paper J1.

21. P. Gong and B. Post, "Experimental Phase Determinations," ACA Meeting (see above), paper M6.

22. J. Nicolosi, "Application of n-Beam Diffraction for Determining the Phases of Certain Reflections," ACA Meeting (see above), paper M7.

23. B. Post, P. Wang and T. Hom, "Intensity Attenuation and Enhancement in 3-beam Diffraction," Z. Naturforsch. 37a, 528 (1982).

24. H. J. Juretschke, "Invariant-Phase Information of X-ray Structure Factors in the Two-Beam Bragg Intensity Near a Three-Beam Point," Phys. Rev. Lett. 48, 1487 (1982).

25. H. J. Juretschke, "Non-centrosymmetric Effects in the Integrated Two-Beam Bragg Intensity Near a Three-Beam Point," Phys. Letters 92A, 183 (1982).

26. H.J. Juretschke, "Modified Two-Beam Description of X-ray Fields and Intensities Near a Three-Beam Diffraction Point, General Theory and First Order Solution" Acta Crystallographica (in press).

27. H.A. Bethe, Ann. Phys. Lpz 87, 55 (1928).

28. Y. Kashiwase, Y. Kainuma, and M. Minoura, *Acta Cryst.* A 38, 390 (1982).

29. B. Post, J. Nicolosi, and J. Ladell, "Experimental Procedures for the Determination of Invariant Phases of Centrosymmetric Crystals." Accepted by Acta Crystallographica.

30. Physics News in 1983, Am. Inst. of Physics (Nov. 1983) pp. 57. Reprinted in Physics Today 37 S21 (Jan. 1984).

31. P.P. Gong and B. Post, "The Experimental Determination of Phases of Reflections from Mosaic Crystals: $ZnWO_4$," Acta Crystallographica A 39, 719 (1983).

32. B. Post, "The Experimental Determination of the Phases of X-ray Reflections," Acta Crystallographica A 39, 711 (1983).

D. SINGLE-LAYER AND MULTILAYER THIN FILMS: THEIR FABRICATION PROCESSES AND THEIR ELECTRONIC, ACOUSTIC, AND OPTICAL PROPERTIES

Professor W.C. Wang and Dr. S. Onishi

Unit SS4-4

1. OBJECTIVE(S)

During the past few years we have developed a novel, superior and cost-effective sputtering facility, around which our program is based. With that new fabrication facility, we have successfully produced certain thin films of excellent quality, and our present program builds on that expertise. We have also begun certain analyses and measurements of basic acoustoelectric effects.

The objectives behind our program fall into two broad, but related, categories.

(i) The larger category involves the development of thin films of the highest quality for acoustoelectric (AE) and acousto-optic (AO) applications. In the selection of film-substrate combinations we will be guided by their potential utility in future monolithic AE or AO devices, although this thin film research should be of benefit well beyond this direct interest. The investigations in this category require two fundamental aspects. The first is that, in the course of developing films of the highest quality, we will need to better understand and to improve the fabrication processes themselves, which are basic to all thin film deposition. The second aspect is that we must perform various analyses and measurements on the electronic, acoustic and optical properties of these thin films, and on the propagation characteristics of the thin film structures that utilize them.

(ii) The related broad category involves basic acoustoelectric processes relating to monolithic SAW devices, and is of interest in itself but is relevant here because the proper understanding of these processes affects the measurement and evaluation of the properties of the thin films under study, and may even influence the selection of optimum film-substrate combinations. The objective here is to analyze various acoustoelectric interactions in thin film structures taking into account the surface and interfacial conditions, and then to experimentally verify the validity of these analyses.

2. APPROACH

Films deposited on various substrates by different means (including our newly developed sputtering system) and under different conditions will be systematically investigated, so that the optimum deposition conditions, such as temperature, pressure, deposition rate and substrate preparation for a certain film on a particular substrate, can be found. Such information will then be translated for application to our own sputtering system, where appropriate. Both conventional thermal and laser annealing will be performed and compared. Recrystallization, thermal stress and composition will be analyzed. Film qualities such as refractive index, acoustic and optic loss, electromechanical coupling coefficient, and different electronic parameters will be measured. These measurements are to be correlated with the orientation, size and density of the grain, surface smoothness, color of the film, etc. Surface and interfacial properties of the thin films will be determined by conventional capacitance measurements as well as by SAW techniques. Special emphasis will be placed upon investigating the thin film qualities of AℓN on SOS (Si on sapphire), InSb on $LiNbO_3$, ZnO on GaAs,etc., since these structures may be useful in certain AE and AO devices. Acoustic propagation characteristics as well as acoustoelectric interactions in these thin film layered structures will be studied both in theory and experiments. In the case of AE interactions in layered structures, a set of field-coupled, two-dimensional differential equations will be analyzed with the involvement of surface and interfacial conditions in the thin film structures.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

1) One of our ongoing programs has involved the fabrication of InSb films of superior quality. The films we have been able to fabricate are probably the best in the world outside of the Soviet Union, but they still require improvement. Toward this goal, we are in the process of installing two new stations in our fabrication facility which are to be used exclusively for the deposition of InSb films and the overlay aluminum nitrate (AℓN) films. An ion beam sputtering station for InSb deposition has been completed and set into operation. The second station for the AℓN films, which are used as the protective overlay for the InSb films, has also been set into operation.

2) In connection with our program on monolithic thin film semiconductor convolvers, we have supplemented our experimental efforts with a theoretical analysis of the transverse component method of operation. The problem is actually three-dimensional and therefore very complicated, but, in the case of InSb on $LiNbO_3$, the InSb layer is sufficiently thin that one can assume that the potential function is independent of the thickness across the InSb film. With that approximation, theoretical calculations have been carried out for InSb on $LiNbO_3$.

3) We have this year, in conjunction with our InSb thin film SAW amplifier program, investigated a new type of ultrasonic imager utilizing multistrip track-changer reflectors. This new type of imager has the advantage over the conventional one of not requiring clock signals. Discussions in the next two sections indicate the structure required and outline the challenges. A single track changer has been successfully constructed.

4) A monolithic thin film real time correlator has been realized in ZnO on Si form and in ZnO on SOS form. The correlator involves two modes of propagation (the Rayleigh mode and the Sezawa mode) in the ZnO film; their distinctly different velocities permit the generation of the correlation signals in a novel manner. Early results from this study have been presented[14] at the October 1983 IEEE Ultrasonics Symposium.

5) For some applications, such as acoustoelectric imagers employing ZnO on glass, the ZnO film must be photosensitive. We have found that the photosensitivity of ZnO can be controlled by different annealing conditions, particularly the annealing temperature, where, for example, appropriate annealing produced a change in photoconductivity of five orders of magnitude.

4. STATE OF THE ART AND PROGRESS DETAILS

During the past few years we have made the following major accomplishments.

(1) We have successfully designed and constructed a novel and superior thin film deposition system utilizing a modified sputter gun. This new system permits us to fabricate thin piezoelectric or semiconducting films of exceptionally high quality at both low temperature and low pressure. The new system has many additional advantages: it is cost effective (low cost both to construct and to operate), it has a high deposition rate (comparable to that of the more costly and less advantageous planar magnetron system), it reduces greatly the problem of thermal stresses, and it permits the fabrication of clear, well-oriented, high quality films on many different substrates, in contrast to other systems.

(2) Using our newly-developed sputtering system, we have fabricated excellent thin films of ZnO and AℓN on various substrates. Good InSb films have also been made using evaporation techniques. The ZnO films, which came first, are as good as any made elsewhere, to our knowledge, and have received high praise from industry. We have also fabricated them on glass, silicon and quartz substrates, with equal facility and quality. The low pressure associated with our sputtering system permits us to produce AℓN films of exceptional quality, colorless and with an extremely smooth surface. The InSb films we have produced are not yet of top quality, but to our knowledge they are the best in the world outside of the Soviet Union, and we have achieved bidirectional SAW amplification with them, which indicates that they are already quite good.

(3) Because of discrepancies which consistently appeared in experimental data involving the acoustoelectric (AE) effect, we theoretically examined the Weinreich relation, which relates the acoustoelectric current to the acoustic power loss and the carrier mobility. A two-dimensional analysis was carried out on a configuration consisting of a semiconductor film on a SAW substrate. It was found that: (i) The Weinreich relation, in SAW applications dealing with semiconductor surfaces, should be modified with the inclusion of trapping dynamics, and (ii) In calculating the d-c acoustoelectric current, the diffusion effect has to be included.

(4) Since monolithic thin film semiconductor convolvers are of substantial interest and should become feasible now that very good thin films can be fabricated, we have succeeded in fabricating an InSb on quartz transverse convolver. This type of convolver with a quartz substrate should provide us with superior temperature stability. A talk is to be presented at the 1984 Ultrasonics Symposium on this subject.

(5) More recently, in conjunction with our InSb thin film SAW amplifier program, we initiated the study of a new type of ultrasonic imager utilizing multistrip track-changer reflectors. This type of scanner has the advantage of operating without the need of clock signals. A single track-changer reflector has been constructed and successfully tested. The goal, of course, is to construct a serpentine scanner utilizing successive track-changer reflectors. The crucial tasks to be achieved are (i) the aperture of the track-changer reflectors has to be small, since the resolution of the imager is inversely proportional to the aperture size, and (ii) the InSb surface acoustic wave amplifier has to be properly incorporated and integrated into the imager, so that the sonic attenuation can be compensated.

(6) A monolithic thin film real time correlator has been realized utilizing the structure of ZnO on Si (also ZnO on SOS). In the ZnO film both the first mode (Rayleigh) and second mode (Sezawa) were excited. The same mask was used for the generation of both modes. The first mode, with the IDT deposited at the interface of ZnO and Si, is operated at a center frequency of 130 MHz, and the second, with the IDT on the ZnO surface, operated at 220 MHz. Thus, their velocities are distinctly different, with a ratio of 1.7 to 1. The correlation output was taken in the width direction transverse to both the sonic propagation and the substrate normal. Transverse DC field was also applied to enhance the output signal. The design of the correlator is such that both the waves are propagating in the same direction. We let the slow mode be launched first; then, when the fast mode overtakes the slow mode, correlation signals are generated via the acousto-electric interaction between the two modes. A talk on this study has been presented at the 1983 IEE Ultrasonics Symposium.[14]

Some of the details regarding our progress and accomplishments are briefly summarized below, together with the state of the art, where relevant.

A. ZnO Films Grown at Low Temperature by Our Newly Developed "Modified Sputter Gun" System[1]

The deposition of C-axis oriented ZnO films by a sputtering technique was initiated in the mid-1960s.[2] Since then, ZnO films have been considered to be one of the most promising materials for optical waveguiding, acoustoelectric applications, acoustooptic interaction media, and SAW transducers, based on its high electromechanical coupling coefficient and its superior optical qualities.[3-7] However, ZnO films did not gain momentum until the group at Kyoto University, Japan, reported several years ago that high quality ZnO films can be obtained on 7059 glass substrates at a relatively high growth rate (≅ 2 μm/hr) by utilizing planar magnetron sputtering.[8-10] (The thermal expansion coefficients of ZnO and 7059 glass are nearly identical.) It is of interest that Dr. S. Onishi, one of the coauthors of this work unit, was a key member of that group under Professor Shiosaki, and that the work on planar magnetron sputtering was the topic of his doctoral dissertation.

However, the films on other substrates, such as silicon, quartz, etc., are normally not optically transparent due to their mismatch in thermal expansion coefficients. In order to improve the quality of ZnO films on all substrates, we studied these past techniques and recognized that in order to achieve clear films one must essentially avoid the problem of thermal stress caused by electron bombardment. We then found how to avoid such stress by suitably modifying the "research S-gun" sputtering system. Presently, this modified S-gun provides us with films of the highest quality. The reason why will become clear as we review the past sputtering methods.

(1) Using a conventional sputtering unit: the substrate holder is of positive polarity. While the ionized A^+ ions strike the targets (ZnO or Zn) to remove material for deposition, the electrons are bombarding the substrate since it is at positive potential. As a result, the substrate temperature is arbitrarily high and can not be controlled. Films thus obtained are of poor quality.

(2) Utilizing a planar magnetron sputtering system: With this system, the majority of the electrons are trapped in the magnetic field. Nevertheless, the substrate will still be bombarded by some of the electrons, since the substrate holder is at positive potential. As stated earlier, the planar magnetron system has significantly improved the film deposition technique. However, ZnO films which are well oriented, clear and of high quality are usually difficult to obtain on other than the 7059 glass substrate.

With the so-called "research S-gun" sputtering system, it was necessary for us to augment and specially arrange the magnetic field so that the thin-film surface is completely outside of the plasma column and free from electron bombardment. The sputtering system is shown in Fig. 1, where an additional magnet is used to shape the plasma column and deflect the electron beam leakage. When these modifications have been made, the S-gun system is the system that has provided us with the best quality of film on various substrates, for the following reasons:

(a) The substrate holder is at neutral potential, and the substrate surface is free from electron bombardment.

(b) The water cooling system is very effective, since it is right next to the target. The target will therefore be at a lower temperature and the substrate surface will not be subjected to excessive heat radiation from the target.

(c) The magnetic field is several times stronger than that associated with a planar magnetron. The plasma column is therefore more tightly confined.

(d) The deposition rate of an S-gun system is comparable to that of a planar magnetron, whose deposition rate is good.

Therefore, the substrate holder is at a low temperature, around 150°C, rather than about 500°C as with the planar magnetron. This feature is

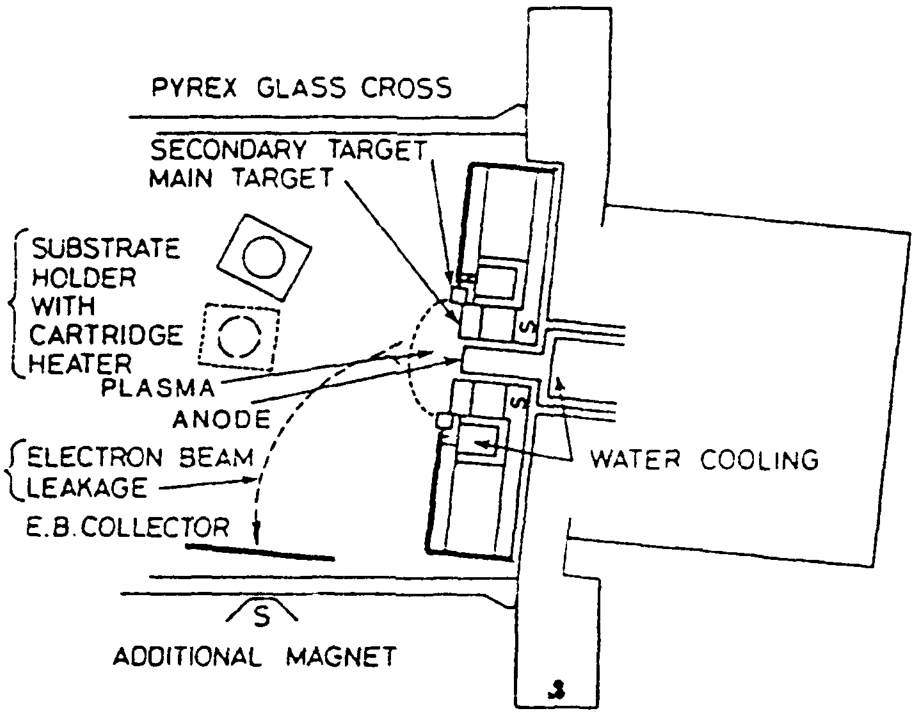


Fig. 1 Geometry of modified sputter gun system.

of key significance, since it greatly reduces the problem of thermal stresses. As a result, with this system, we are able to obtain clear, well-oriented, high quality ZnO films on all substrates available to us, such as glass, silicon and quartz substrates.

In addition to the key features of low pressure and low temperature sputtering, other advantages of this newly-developed sputter gun system include: (a) a much larger area of high quality thin film can be obtained, and (b) the system is cheaper to construct and cheaper to run.

In certain applications, such as ZO on Si acoustoelectric devices, the ZO film is required to be of very high resistivity and highly transparent, and in other cases, such as ZO on glass acoustoelectric imagers, the ZO film should be photosensitive. We have found that the ZO photosensitivity can be controlled by different annealing conditions, especially the annealing temperature. For instance, a 5 orders of magnitude change in photoconductivity (between dark and UV light) has been achieved for samples under an annealing temperature $\approx$ 100°C and an annealing duration of about 5 minutes.

B. AℓN Films Grown at Low Pressure and Low Temperature, Utilizing Our "Modified Sputter Gun" System

The utilization of our modified sputter gun has also been extended for the deposition of aluminum nitride (AℓN) films on various substrates.[11-14] The strong magnetic field associated with this gun system makes it possible to maintain a stable plasma at a total sputtering gas pressure (TSGP) as low as 1 μTorr. This low-pressure sputtering provides films with denser and finer grains and it improves c-orientation, surface smoothness, film color, transparency and deposition rate. Our experience shows that the qualities of AℓN films are extremely sensitive to the level of total sputtering gas pressure, the lower the better. The total pressure in our system varies from 1 μTorr to 8 μTorr. The deposition rate of the sputter gun is three times as high as that of a conventional planar magnetron, since the target cooling system makes it possible to increase rf or dc power. It is 2.4 μm/hr at an input power of 300 watts. The experimental results in Fig. 2 show that the growth rate decreases with increase in total pressure. And the film is only transparent and colorless at low pressure sputtering. High quality films are deposited at temperatures around 350°C, which is low for AℓN deposition. The substrate surface in this system, as stated in A, is free from charged particle bombardment and excessive heat radiation.

The acoustic properties of the AℓN films have been revealed by evaluating delay line characteristics.

(i) The insertion loss of an AℓN on SOS delay line is about 20 dB at a center frequency of 250 Mhz, which indicates that the film is of excellent quality and that the piezoelectrical coupling coefficient approaches that of a single crystal. The design parameters for the delay line are:

propagation distance, 1 cm
IDT finger pairs, 12
IDT aperture, 0.1 cm
AℓN film thickness, 4 μm
Si film thickness, 1 μm

(ii) Dispersion characteristics: Velocity dispersion characteristics for the AℓN-Si-sapphire structure have been obtained from delay lines fabricated with different AℓN film thicknesses. As expected, the characteristics of velocity dispersion fall between those of AℓN and those of sapphire.

C. InSb Films

The deposition of InSb thin films on $LiNbO_3$ substrates has been accomplished at different centers since the early 1970s,[15 18] and the process has been used with little variation in Japan, France, Russia and the U.S. However, due to difficulties in obtaining high quality films, acoustoelectric devices using InSb films have not been proven to be practical. For AE applications it is required that the InSb film possess high drift mobility and high resistance. The best data reported to date are by Kotelyanskii, et al.,[18] in Russia; they can obtain a film of 500Å thickness and a drift mobility of 1600 cm^2/v-sec. This mobility is over three times better than what has been reported in the

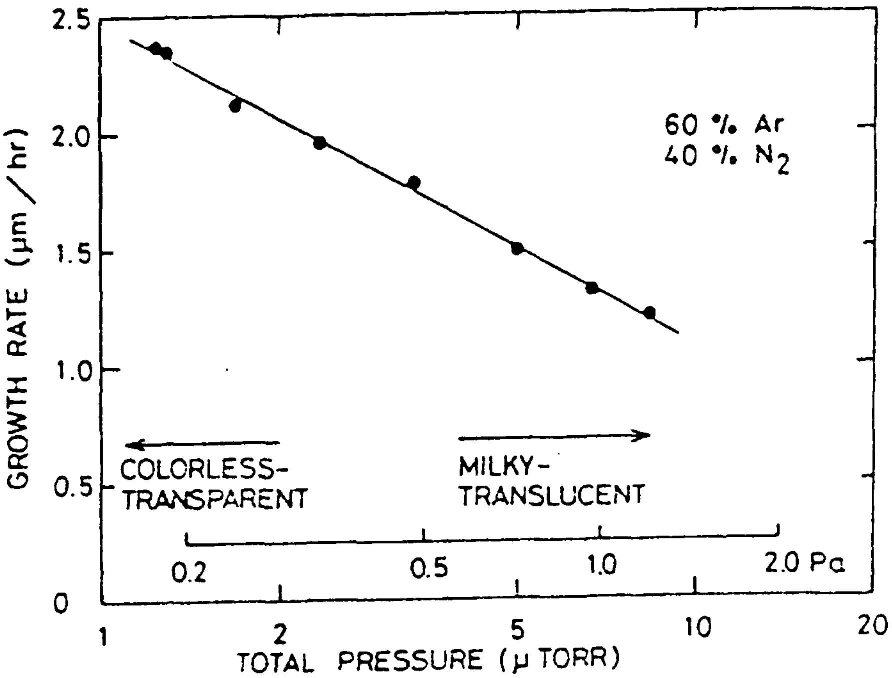


Fig. 2 Deposition condition of AℓN film: total sputter gas pressure vs. growth rate.

western world. We have been engaging in the development of InSb films for three years and have achieved bidirectional SAW amplification with those films necessary for reducing the insertion loss of a convolver.[19] The drift mobility we obtained here is ≅ 700 cm^2/v-sec at a film thickness ~ 500Å, which is better than what is reported in this country but is still low in comparison with that of Kotelyanskii, et al.[18] Private discussions with Kotelyanskii during his visit to us here revealed that their set-up is much more elaborate and costly, and they have invested 10 years in this effort.

Figure 3 shows some of our experimental results of Hall mobility vs. the conductivity thickness product. The film thickness is about 450Å. Our process, especially on forming the electric contacts, we feel is superior.[19] The highest Hall mobility obtained by us is about 2500 cm/volt-sec, which means that if the traps can be eliminated in this film a drift mobility of that same magnitude can be achieved. The significance of these measurements is that they tell us how good a film can be achieved ultimately, when the Hall and drift mobilities are equal. Thus, these measurements offer us substantial hope.

We have reason to believe that our InSb film quality will be greatly improved, since we have installed two new stations which are used exclusively for the deposition of InSb films and its overlay aluminum nitrate (AℓN) film. The ion beam sputtering station for InSb deposition has been completed and is now set into operation. The quality of the

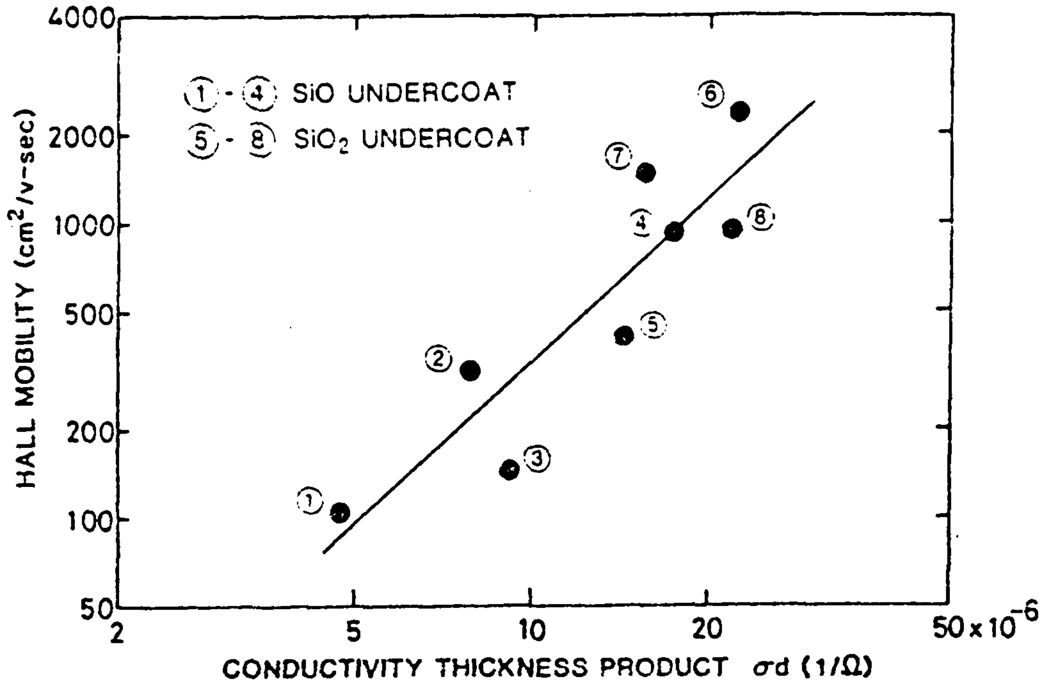


Fig. 3 Hall mobility of InSb film (450Å thick) versus conductivity thickness product.

film after a period of adjustment and experimentation is expected to be superior to previously produced films. The second station for AℓN films used as a protective overlay for the InSb film has also been set into operation. Excellent AℓN films have been produced by the new station.

D. Examining the dc Acoustoelectric Current Produced by SAW

It is generally believed that surface acoustic waves will be useful in determining the surface and interfacial properties of semiconductor wafers and semiconductor films; however, in the process of our correlating data from SAW measurements with those obtained from conventional methods, some discrepancies consistently appear.[20-24] This led us to examine the validity of the commonly-used Weinreich relation for a configuration consisting of a semiconductor on a SAW substrate. The Weinreich formula relates the acoustoelectric current to the acoustic power loss and the carrier mobility. Since both the AE current and the power loss are measurable quantities, the carrier mobility can thus be determined. Comparing the mobility measured here with the Hall mobility measurement, trapping dynamics can also be revealed.

The Weinreich relation was first derived for phonon-drag phenomena in Ge due to the deformation potential.[25] Its validity in bulk piezoelectric semiconductors was tested experimently by one of us (W.C. Wang) in 1962. We found that the relation between attenuation constant (gain) and acousto-electric current does not obey Weinreich's formula, which predicts zero current at zero gain. Subsequently, H.N. Spector pointed

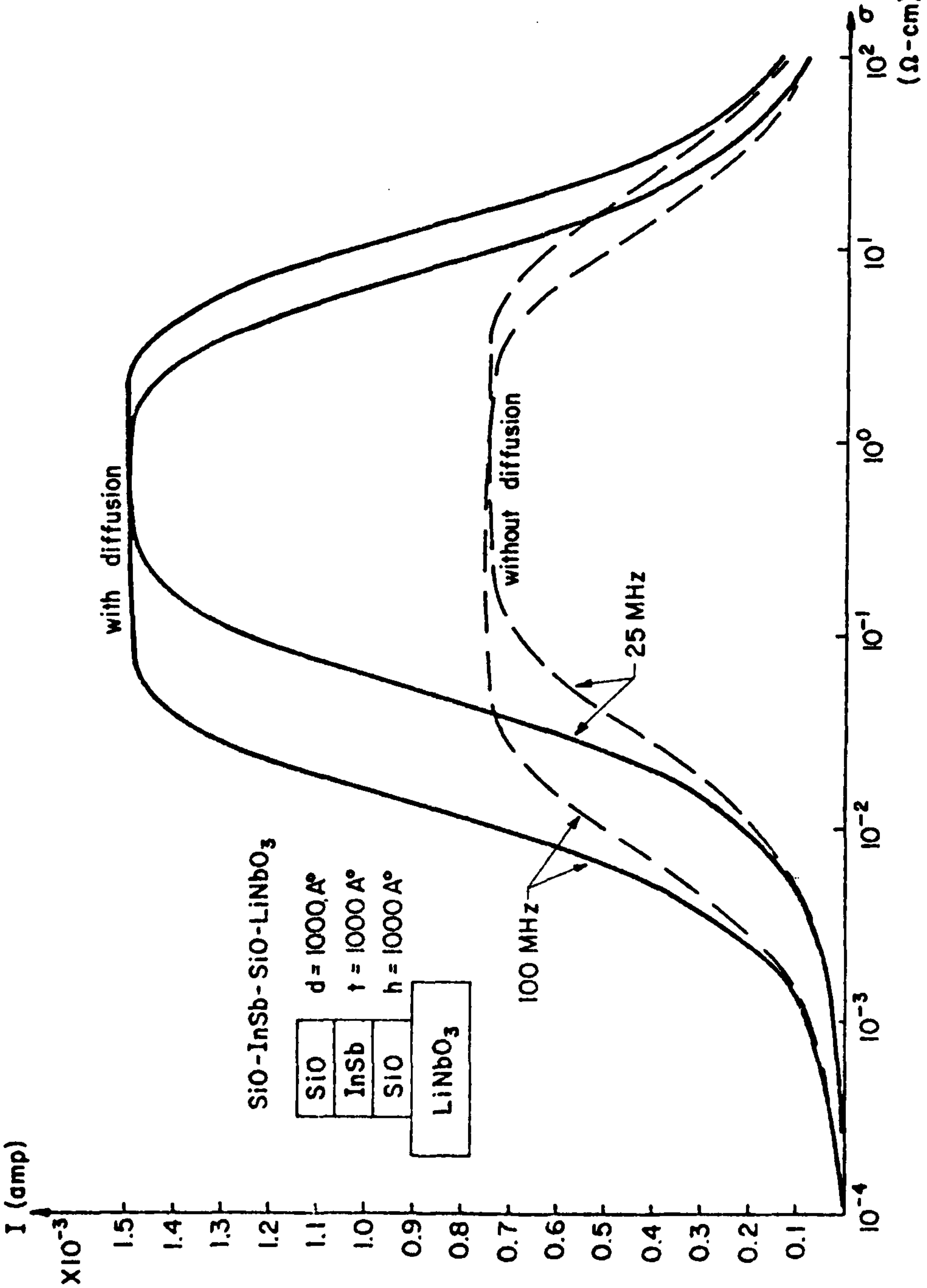


Fig. 4 Short circuited dc acoustoelectric current as a function of conductivity.

out that the inconsistency between theory and experiment was due to the presence of traps, and they properly modified Weinreich's formula with the inclusion of trapping dynamics. It is further found that, in SAW applications, especially when dealing with semiconductor surfaces, not only must the trapping dynamics be considered but the diffusion effect should be included as well. Figure 4 is a typical curve which shows the differences obtained between the short-circuited dc acousto-electric currents when calculated with and without the diffusion effect. It is noted immediately that the diffusion current cannot be ignored.

E. Monolithic Thin Film Semiconductor Convolver[29-31]

On the development of the monolithic thin-film semiconductor convolver, we have progressed well both experimentally and theoretically. In experiment, we have succeeded in demonstrating the operation of an InSb on quartz substrata convolver. Due to the weak electromechanical coupling of quartz, the operation of the quartz substrate convolver has not been realized until now. Quartz substrate convolvers are expected to offer superior temperature stability, which has been the major draw-back of $LiNbO_3$ substrate convolvers. The detailed characteristics of the quartz convolver, such as dynamic range, insertion loss, dispersion as well as temperature stability, are under investigation. The configuration of the InSb on quartz convolver is shown in Fig. 5(a).

An InSb film ~500Å thick is deposited on the ST cut (also on x-cut, y-propagating) quartz. The IDT transducer aperture is 0.93 mm wide. The convolver output is taken in the x-direction, transverse to the sonic propagation, but in the plane of the substrate surface. A dc biasing voltage is applied in the transverse direction to enhance the convolution output. Without dc biasing, no convolver output is observed. The magnitude of convolver output depends on the strength of the dc biasing field and the quality of InSb film.

Theoretical calculations have been carried out. Good agreement has been obtained between theory and experiment for the case of an InSb on $LiNbO_3$ substrate. In the case of quartz substrates, investigation is still in progress. Approximations have been used in the theoretical derivation. Strictly speaking, the problem is a three-dimensional one. However, since the thickness of the InSb film is only a fraction of the Debye screening length, the potential function is essentially constant across the thickness of the film. With the assumption of constant potential, the problem reduces to a two-dimensional one.

The analytic model for the transverse convolver is indicated in Fig. 5(b), which shows the top view of the transverse convolver. By writing down the conduction current density for the InSb film, combined with the continuity equation and Gauss's law, and by utilizing an electrostatic approximation, a fourth-order differential equation in potential is obtained. The following briefly outlines the procedure for solving the equation. The boundary conditions for open-circuit voltage at frequency ω are:

(i) Potential $\phi(d/2) = \phi(-d/2)$, when no dc biasing voltage is applied.

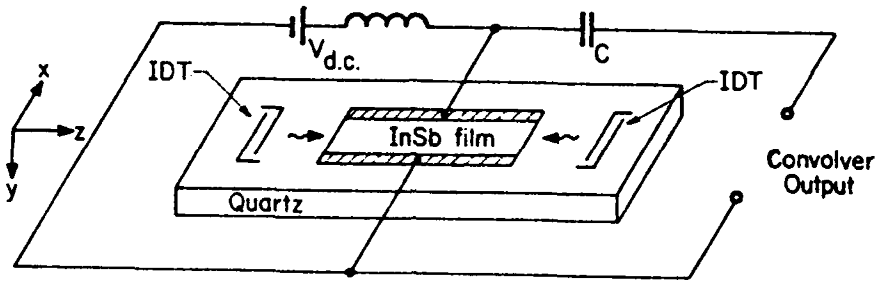


Fig. 5(a)

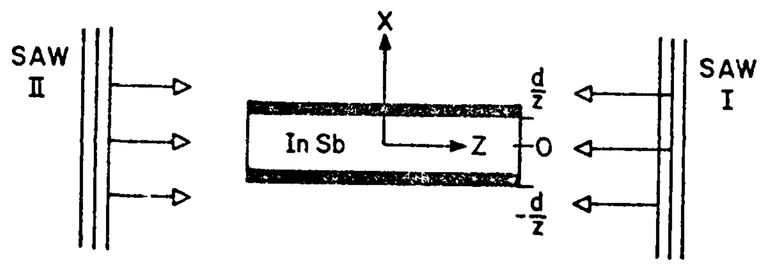


Fig. 5(b)

Fig. 5(a,b) Configuration of an InSb on quartz transverse convolver.

(ii) Current density $J(d/2) = J(-d/2) = 0$

(iii) Total power across the width of the InSb film is constant.

With this set of boundary conditions one will obtain the expressions for the nonlinear correct source ($n_\omega^- E_\omega^+$) due to two oppositely propagating SAWs, which is the driving source for convolver signal generation. The boundary conditions for solving the open circuit voltage at 2ω (convolver output) are:

(i) $J_{2\omega} = 0$ at $x = -d/2$,

(ii) $J_{2\omega} = 0$ at $x = d/2$,

(iii) $\phi(d/2) = \phi(-d/2)$, when no dc biasing voltage is applied.

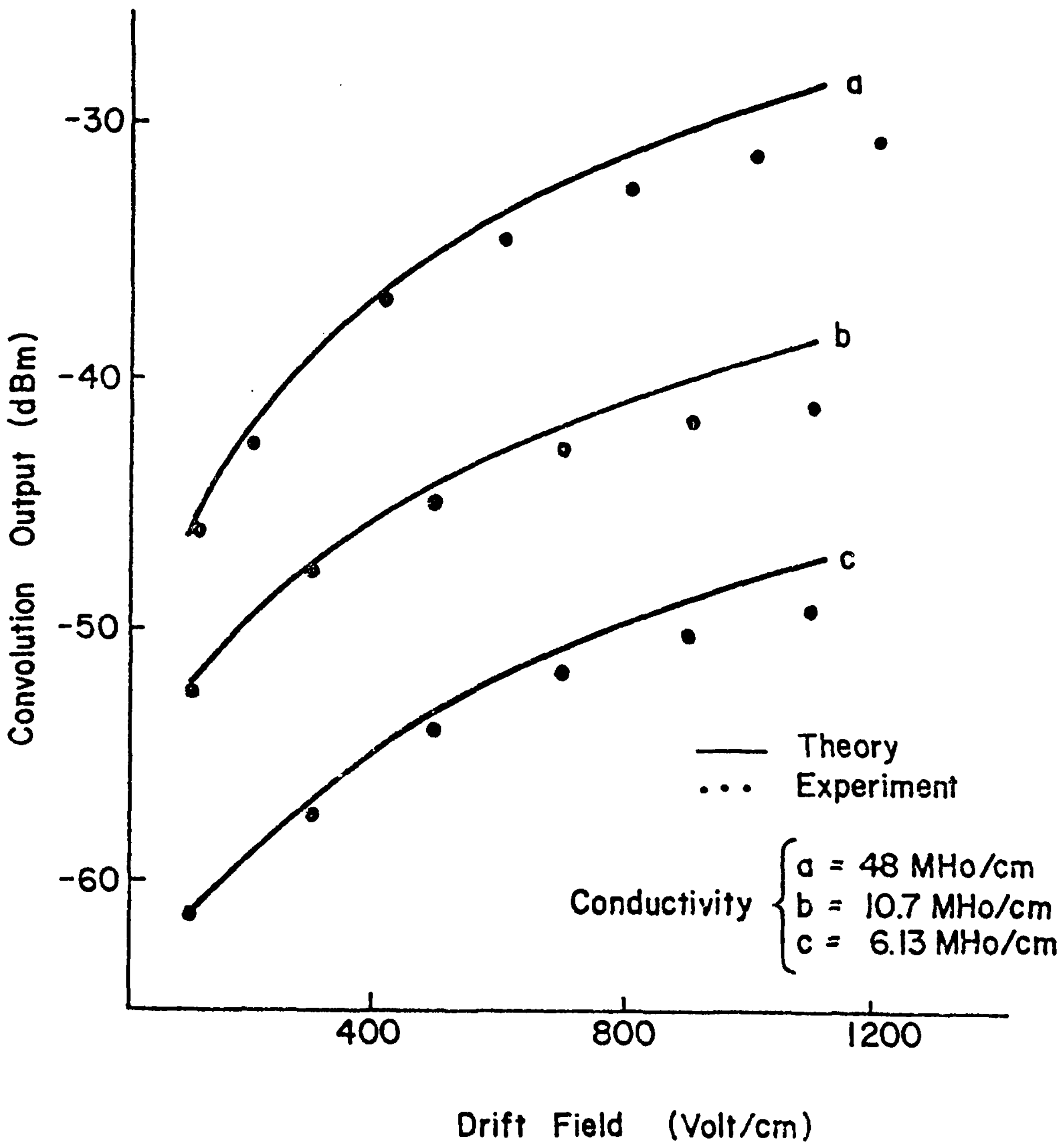


Fig. 6 Convolver output as a function of transverse biasing field.

In the process of solving the differential equation, one further assumes that the potential function at ω is in the form $\phi(x)e^{j(kz\pm\omega t)}$ and at 2ω in the form $\phi(x)e^{j2\omega t}$. Finally, we obtain the convolution output as a function of the dc biasing field with conductivity as a parameter. a typical set of curves is shown in Fig. 6.

It should be added that industry has expressed substantial interest in our thin film fabrication capability and its potential for producing excellent films in a cost-effective manner. For example, we have a contract with Fairchild Camera and Instrument for a program that would utilize an InSb thin film amplifier in the construction of a compact and low-loss SAW delay line for repeater-jammer applications.

5. REFERENCES

1. S. Onishi, M. Eschwei and W.C. Wang, Appl. Phys. Lett. 38, 419 (1981).
2. S. Wanuga, T.A. Midford and J.P. Dietz, IEEE Ultrasonics Symposium (1965).
3. F.S. Hickernell, J. Appl. Phys. 44, 1061 (1973).
4. F.S. Hickernell and J.W. Brewer, Appl. Phys. Lett. 21, 389 (1972).
5. R.S. Wagers, G.S. Kino, P. Galle and D. Winslow, IEEE Ultrasonics Symposium (1972).
6. R.F. Pierret, R.L. Gunshor and M.E. Cornell, IEEE Ultrasonics Symposium (1979).
7. S. Furukawa, S. Tsuzihara, T. Moriizumi and T. Yasuda, IEEE Ultrasonics Symposium (1979).
8. T. Shiosaki, S. Onishi and A. Kawabata, J. Appl. Phys., 50, 3113 (1979).
9. T. Onishi, Y. Hirokawa, T. Shiosaki and A. Kawabata, Japanese J. of Appl. Phys. 17, 773 (1978).
10. F.S. Hickernell, IEEE Ultrasonics Symposium (1980).
11. S. Onishi, M. Eschwei, S. Bielaczy and W.C. Wang,to appear in Appl. Phys. Letters (October 15, 1981).
12. A.J. Noreika, M.H. Francombe and S.A. Zeitman, J. Vac. Sci. Technol. 6, 194 (1969).
13. A.J. Shuskus, T.M. Reeder and E.L. Paradis, Appl. Phys. Lett. 24, 155 (1974).
14. T. Tsubouchi, K. Sugai and N. Mikoshiba, Proc. IEEE Ultrasonic Symp. 446 (1980).

15. L.A. Coldren, Appl. 18, 319 (1971).

16. L.A. Coldren and G.S. Kino, Appl. Phys. Lett. 23, 117 (1973).

17. K. Yamanouchi, K. Abe and K. Shibayama, Suppl. J. Japan Soc. Appl. Phys. 43, 203 (1974).

18. L. M. Kotelyanskii, A. I. Krikunov, A. V. Medved, R.A. Mishkinis and V.V. Panteleev, Sov. Phys. Semicond. 12, 751 (1978).

19. M. Eschwei, S. Onishi, F. Cassara and W.C. Wang, to be presented at American Vacuum Society, 18th National Vacuum Symposium (1981).

20. K.S. Chen, Z.S. Wu, H. Schachter and W.C. Wang, to be presented at IEEE Ultrasonics Symposium (1981).

21. J.H. Cafarella, A. Bers and B.E. Burke, IEEE Symposium Proceedings, 181 (1972).

22. A. Bers, J.H. Cafarella and B.E. Burke, Appl. Phys. Lett. 22, 399 (1973).

23. R. Adler, D. Jones, S. Datta and B.J. Hunsinger, IEEE Ultrasonics Symposium (1980).

24. P. Das, R.T. Webster, H. Estrada-Vazquez and W.C. Wang, Surface Science 86, 848 (1979).

25. G. Weinreich, Phys. Rev. 107, 317 (1957).

26. W.C. Wang, Phys. Rev. Lett. 9, 443 (1962).

27. H.N. Spector, J. Appl. Phys. 34, 3628 (1963).

28. K.A. Ingebrigtsen, J. Appl. Phys. 41, 454 (1970).

29. W.C. Wang and P. Das, IEEE Ultrasonics Symposium (1972).

    G.S. Kino, W.R. Shreve and H.R. Gautier, IEEE Ultrasonics Symposium (1972).

    J.M. Smith, E. Stern, A. Bers and J. Cafarella, IEEE Ultrasonics Symposium (1973).

30. L.P. Solie, Proc. IEEE 64, 760 (1976).

31. P. Defranould and C. Maerfeld, Proc. IEEE 64, 748 (1976).

6. PUBLICATIONS

1. L. Rosenheck, H. Schachter and W.C. Wang, "Acoustoelectric Real Time Correlator," Japanese Journal of Applied Physics, Vol. Supp. 18-1, p. 215 (1979).

2. P. Das, R. Webster, H. Estrada and W.C. Wang, "Contactless Semiconductor Surface Characterization Using Surface Acoustic Waves," Surface Science, Vol. 86, p. 848 (1979).

3. W.C. Wang, H. Schachter and F. Cassara, "Acousto-Electric Surface Acoustic Wave Demodulators," Proc. IEEE Ultrasonics Symposium, Cat. #79-CH-1482-9SU (1979).

4. S. Onishi, M. Eschwei and W.C. Wang, "Transparent and Highly Oriented ZnO Films Grown at Low Temperature by Sputtering with a Modified Sputter Gun," Applied Phys. Lett., Vol. 38(6), p. 319 (1981).

5. S. Onishi, M. Eschwei, S. Bielaczy, and W.C. Wang, "Colorless-Transparent, c-Oriented Aluminum Nitride Films Grown at Low Temperature by a Modified Sputter Gun," Applied Phys. Lett., 39, 643 (1981).

6. M. Eschwei, S. Onishi and W.C. Wang, "Properties of InSb Film for SAW Amplifier," accepted for presentation in American Vacuum Society 28th National Vacuum Symposium (1981).

7. Z.S. Wu, S. Onishi, K.S. Chen and W.C. Wang, "Monolithic Thin Film Acoustoelectric Convolver," presented in 1981 Ultrasonics Symposium.

8. S. Onishi, M. Eschwei, S. Bielaczy and W.C. Wang, "Acoustic, Optical Properties of AℓN Films Grown at Low Temperature by Modified Sputter Gun," presented in 1981 Ultrasonics Symposium.

9. K.S. Chen, Z.S. Wu, H. Schachter and W.C. Wang, "D-C Acoustoelectric Current Induced in a Semiconductor Film Adjacent to a SAW Substrate," presented in 1981 Ultrasonics Symposium.

10. M. Eschwei, S. Onishi, F. Cassara and W.C. Wang, J. Vac. Sci. Technol., 20, 873 (1982).

11. S. Onishi, M. Eschwei, B. El-Asir, S. Bielaczy, H. Schachter and W-C. Wang, "Enhancement of Monolithic Convolver Output by Transverse Biasing," presented at IEEE Ultrasonic Symposium (October 1982).

12. S. Onishi, M. Eschwei, B. El-Asir and W-C. Wang, "Enhancement of a Thin Film Indium Antimonide on $LiNbO_3$ Convolver by Transverse Biasing," Japanese J. of Appl. Phys., 22, 273, (May 1983).

13. F.A. Cassara and W-C. Wang, "Laboratory Experiments in SAW Devices and their Applications," IEEE Trans. on Ed., 26, 52 (May 1983).

14. S. Onishi, B. El-Asir, M. Eschwei and W-C. Wang, "Monolithic Thin Film Acousto-Electric Correlator Utilizing Two Modes of SAW," presented at IEEE Ultrasonics Symposium (October 1983).

SECTION III
INFORMATION ELECTRONICS

SECTION III: INFORMATION ELECTRONICS

A. ADAPTIVE FILTERING AND SPECTRAL ESTIMATION

Professor A. Papoulis

Unit IE4-1

1. OBJECTIVE(S)

Our objective is the investigation of various theoretical and applied aspects of spectral estimation. This topic is of fundamental interest in many areas, including image processing, speech processing, underwater detection, radar, interferometry, surveillance, system identification, and crystallography. It involves the determination of the spectral properties of one- and two-dimensional, discrete and continuous-time, stationary and nonstationary processes.

2. APPROACH

In the study of adaptive filters, we concentrate on the determination of the statistical properties of the Widrow filter designed either directly in the time domain or in terms of running transforms, and we investigate the effect of round-off errors, stability considerations involving periodic sources, and conditions for asymptotic equilibrium. In the solution, we use an adaptation of Floquet theory to recursion equations and perturbation techniques involving small departures from equilibrium.

In the extrapolation problem, we reexamine the bandlimited extrapolation method developed earlier[13,14] assuming that the data are determined from a single sample of the process. We further determine the class spectra that are consistent with the constraints and we select a particular form that meets prior information and leads to simple computations involving ARMA models.

We extend the method of maximum entropy to problems where the data involve the values of a single sample of the unknown spectrum and we seek to determine the asymptotic properties of the resulting spectra as the length of the sample increases.

We note below selectively some of the approaches. Details of the early results are given in the cited papers.

The problem of hidden periodicities has been considered by Pisarenko[30] and others in the context of a result due to Caratheodory. We employ a method that utilizes Levinson's algorithm and the properties of the prediction error filter. In the method of maximum entropy, we assume that the autocorrelation $R[m]$ of the process $s[n]$ is known for $|m| \leq N$ and that $S(\omega)$ is an all-pole function:

$$S(\omega) = \frac{1_N}{\left| 1 - a_1^N e^{-j\omega T} \cdots - a_N^N e^{-jN\omega T} \right|^2} \qquad (1)$$

The unknown parameters are determined recursively (Levinson):

$$e_{N-1}\Gamma_N = R[N] - \sum_{k=1}^{N-1} a_k^{N-1} R\,[N-k] \qquad (2)$$

$$a_k^N = a_k^{N-1} - \Gamma_N\, a_{N-k}^{N-1} \qquad 1 \leq k \leq N-1$$

$$(3)$$

$$a_N^N = \Gamma_N \qquad e_N = (1-\Gamma_N^2)e_{N-1}$$

In our approach, we assume that the unknown spectrum consists of lines

$$S(\omega) = \lambda_o + 2\pi \sum_{i=1}^{N} \gamma_i \delta(\omega-\omega_i) \qquad (4)$$

and we show that their parameters can be determined by a modification of the above algorithm. We, thus, establish Caratheodory's results using the properties of the error filter

$$E_N(z) = 1-a_1^N z^{-1} - \cdots - a_N^N z^{-N} \qquad (5)$$

This approach is computationally simpler and it leads to a method for estimating the parameters λ_o, γ_i, and ω_i.

In the extrapolation problem, we modify the so-called Papoulis-Gerchberg algorithm[13,14] to problems involving noise and prior energy constraints.

In the echo-cancelling problem (telephone communications) we observe the sum

$$r[n] = s[n] + y[n]$$

where $s[n]$ is the signal coming from location B and $y[n]$ is an attenuated and distorted version of a signal coming from location A. The purpose of the filter is to remove $y[n]$. It has been observed that, if the signal $s[n]$ contains periodicities, the adaptation might diverge. We are investigating of the statistical properties of the filter parameters and the conditions for convergence of the adaptation.

3. SUMMARY OF RECENT PROGRESS

We have considered the general properties of extrapolating spectra under various constraints and we developed new results in the estimation of spectra of FM signals and in the determination of the maximum entropy solution under nonuniform data constraints. We list below the principal publications and we follow with a detailed description of some of these results.

SECTION III: INFORMATION ELECTRONICS

List of Recent JSEP Publications

1. A. Papoulis, "Predictable Processes and Wold's Decomposition," IEEE, ASSP, submitted.

2. A. Papoulis, "Levinson's Algorithm, Wold's Decomposition, Spectral Estimation," SIAM Journal, to appear March 1985.

3. A. Papoulis, "Innovations, Wold's Decomposition, Lattice Filters, and Spectral Estimation," IEEE Modern Spectral Estimation Seminar, University of Pennsylvania, April 1984.

4. A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw Hill, New York, 2nd edition, 1984. The book contains original material developed under JSEP sponsorship.

5. W. Xu and C. Chamzas, "On the Extrapolation of Band Limited Functions with Constraints," IEEE Trans. Acoustics and Signal Processing, October 1983.

6. A. Papoulis, "On Entropy Rate," Third Workshop on Maximum Entropy and Bayesian Methods, University of Wyoming, August 1-4, 1983. Invited Paper.

7. A. Papoulis, "From Levinson's Algorithm to Wold's Decomposition," SIAM 1983 National Meeting, Denver, Colorado, June 6-8, 1983. Invited Paper.

8. A. Papoulis, "Random Modulation," IEEE Trans. on Acoustics, Speech, and Signal Processing," ASSP-31, pp. 96-105, February 1983.

9. A. Papoulis, "Spectra of FM Signals," Proceedings of the 9th Prague Conference on Information Theory, February 1983.

10. A. Papoulis, "Detection of Line Spectra and Point Sources," Conference on Signal Recovery with Incomplete Information, sponsored by Optical Society of America and Supported by U.S. Air Force Office of Scientific Research, Lake Tahoe, January 1983. Invited Paper.

Spectra of FM Signals

We have investigated the spectral properties of the FM signal

$$x(t) = \cos[\omega_0 t + \phi(t)] \quad \phi(t) = \int_0^t c(\alpha)d\alpha \tag{6}$$

under the assumption that the instantaneous frequency c(t) is a stationary stochastic process. This problem is important in modulation theory[23] and it has applications also in physics.[24] We present below briefly some of our results,[25,26] with

$$w(t) = e^{j\phi(t)} \quad R_x(\tau) = \frac{1}{2} \operatorname{Re} w(\tau)e^{j\omega_0 \tau} \tag{7}$$

It suffices, therefore, to consider only the complex envelope w(t) of the process x(t).

From the assumed stationarity it follows that

$$R_w(\tau) = E\{w(t+\tau)w^*(t)\}$$

$$= E\{\exp \int_t^{t+\tau} c(\alpha)d\alpha\} = E\{w(\tau)\} \tag{8}$$

This shows that the spectral properties of x(t) can be expressed in terms of the mean of w(t). We shall determine them for the following important cases:

(a) We assume that the instantaneous frequency c(s) is a Markoff chain taking the values of α_i with probabilities

$$p_i = P\{c(t) = a_i\} \tag{9}$$

and transition probabilities

$$\pi_{ij}(\tau) = P\{c(t+\tau) = a_j | c(t) = a_i\} \tag{10}$$

With

$$R_{ik}(\tau) = E\{w(\tau) | c(0) = a_i,\ c(\tau) = a_k\}\pi_{ik}(\tau) \tag{11}$$

it follows that

$$R_w(\tau) = \sum_{i,k} p_i R_{ik}(\tau) \tag{12}$$

To determine the autocorrelation of w(t), it suffices, therefore, to find the conditional correlations $R_{ik}(\tau)$ defined as above. To do so, we show that

$$R_{ik}(\tau+\upsilon) = \sum_m R_{im}(\tau)R_{mk}(\upsilon) \tag{13}$$

From this it follows by differentiation that $R_{ik}(\tau)$ satisfies the Kolmogoroff equation

$$R'_{ik}(\tau) = \sum_m R_{im}(\tau)R_{mk}'(0^+) \quad \tau > 0 \tag{14}$$

with the initial conditions

$$R_{ik}(0^+) = \begin{Bmatrix} 1 & i=k \\ 0 & i\neq k \end{Bmatrix} \qquad R'_{in}(0^+) = \begin{Bmatrix} ja_i+\lambda_{ii} & i=k \\ \lambda_{ik} & i\neq k \end{Bmatrix}$$

where $\lambda_{ik}=\pi'_{ik}(0^+)$ are the transition probability rates of the process c(t). Thus, to determine $R_w(\tau)$ we must solve the above system of linear equations. Transforming R_{ik} and inserting into (12), we obtain the spectrum $S_\omega(\omega)$ of w(t). From (13) it follows that *if c(t) has finitely many states, then its spectrum is rational*.

We have applied the above to a number of special cases.[25]

(b) We now assume that

$$c(t) = b(t - \theta) \tag{15}$$

where b(t) is a cyclostationary process with period T and θ is a random variable uniform in the interval (O,T). This is a realistic model in frequency-shift-keying modulation (known frequency statistics) and phase-shift-keying modulation (known phase statistics).

From the above and (8) it follows that c(t) is a stationary process with autocorrelation

$$E\{w(\tau)\} = E\{E\{w(\tau)|\theta\}\} = \frac{1}{T}\int_0^T E\{w(\tau)|\theta\}d\theta \tag{16}$$

To determine the spectrum of w(t), it suffices therefore, E{w(t)|θ}. This problem arises in frequency-shift-keying (known frequency statistics) and in phase-shift-keying (known phase statistics). We have carried out the computations for a number of special cases (FSK, PSK, and DPSK) involving piecewise linear phase modulation.[26] Three typical examples are shown in Fig. 1.

The general class of extrapolating spectra. We have developed the properties of four-terminal lattices and developed a one-to-one correspondence between an arbitrary set of reflection coefficients, a Hurwitz polynomial, and a positive definite sequence. Using these results, we have given a complete characterization of the class of positive sequences starting from given consecutive data

$$R_0,\ R_1, \ldots,\ R_N \tag{17}$$

and we have shown that the corresponding class of extrapolating spectra is a function of the form

$$S(e^{j\omega}) = \frac{P_N[1 - |p(e^{j\omega})|^2]}{|E_N(e^{j\omega}) + e^{-jN\omega}E_N(e^{-j\omega})p(e^{j\omega})|^2} \tag{18}$$

where P_N is the MS error in the estimate of the process $\underset{\sim}{x}[n]$ in terms of its N most recent point values, $E_N(z)$ is the corresponding error filter, and p(z) is an arbitrary reflection coefficient. This basic relationship can also be derived using frequency domain ideas and transmission line theory.[32]

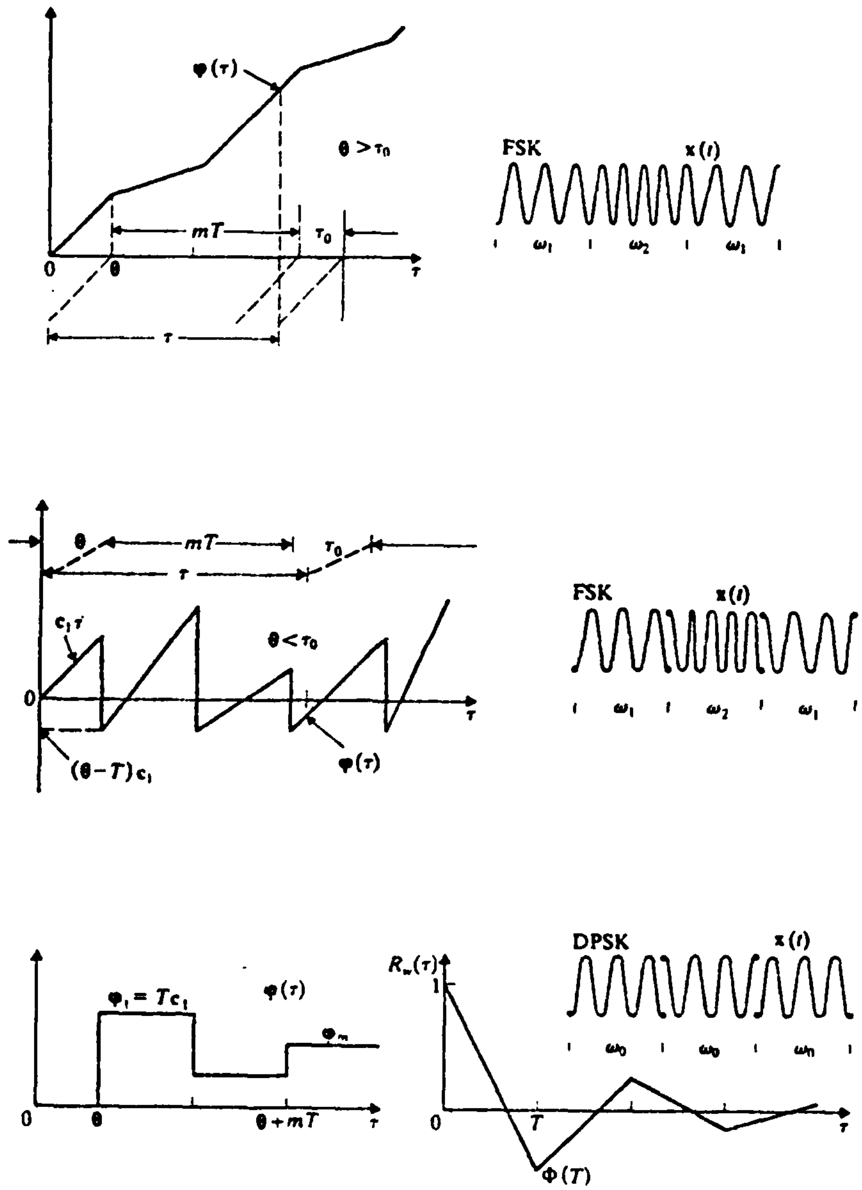

φ(τ)
θ > τ0
mT
τ0
0
θ
τ
FSK
x(t)
ω1
ω2
ω1
θ
mT
τ0
τ
c1τ
θ < τ0
0
(θ−T)c1
φ(τ)
FSK
x(t)
ω1
ω2
ω1
φ1 = Tc1
φ(τ)
φm
0
θ
θ+mT
τ
Rw(τ)
1
0
T
Φ(T)
DPSK
x(t)

Fig. 1

4. STATE OF THE ART AND PROGRESS DETAILS

Most current investigations deal with parametric extrapolation techniques and adaptive methods. The Blackman-Tukey method, the subject of extensive earlier studies,[1] is applied mainly to problems involving large sample sizes. The window problem[2] is still of some interest, but only for specific cases. We would like to note that in an earlier investigation, we developed an optimum class of windows,[3] and used it in image processing[4] and ultrasonics.[5] The parametric extrapolation method started with bandlimited extrapolation techniques based on expansions into prolate spheroidal functions.[6,7] Now, however, it is rarely used. In a more recent study, we developed a new extrapolation method[8] based on an iteration using FFTs, and we applied it in a variety of problems including image restoration[10] and detection of hidden periodicities.[11] The method in its original form is slowly convergent and, for noisy signals, only marginally useful (the problem is ill posed). However, in the presence of prior information[11] or imposed constraints[12] it leads to satisfactory results. More recently, a variety of papers have appeared involving refinements of the method.[13]

In the last years, the method of maximum entropy (MEM)[14,15] has been discussed in a large number of papers. Various generalizations involve ARMA extrapolation[16] and extensions of Burg's method based on the concept of cross-entropy.[17,18] The method is popular because, unlike the method of windows, it can detect lines even if the data length is short. Furthermore, unlike other parametric methods, it involves only the solution of linear equations. In fact, the unknown spectrum can be determined recursively (Levinson's algorithm) because, as it was recognized earlier, the method is equivalent to the solution of the prediction problem. In addition, in the presence of uncertainty, it is optimum in the sense of maximizing the entropy rate of the given process. This argument is used by Jaynes[19] to solve a variety of ill-posed inverse problems (see also B. deFinetti).[20] Recently, the method is used also for large molecular crystallography.[21] In a recent article,[22] we summarize the underlying concepts in the context of spectral estimations.

We are now concentrating on the stochastic version of the spectral estimation problem. This problem is more complex, involving not only the design of desirable estimators but also the determination of their statistical properties.

We are considering various theoretical and applied aspects of this problem with particular emphasis on adaptive techniques. In this context, we examine the possibility of data reduction based on suitable data transformation, focusing on the use of running transforms. We hope to show that this approach reduces the variance of the estimates and simplifies the implementation. We elaborate below on certain aspects of this study.

A technique of spectra estimation, used extensively in many applications, is the method of maximum entropy (MEM). Most problems solved with this method, however, are based on the assumption that the given data form a set of $N + 1$ consecutive integers as in (17). In a number of applications, particularly involving two-dimensional problems, the data are not consecutive. They belong to an arbitrary set D

$$R_m \qquad m \in D \qquad (19)$$

and the problem is to find the MEM spectra satisfying these constraints. Unlike the uniform-data case, this problem does not have a closed form solution. We propose to develop numerical techniques for solving it and ultimately, we seek to characterize the general class of spectra that meet these constraints. We conclude with a brief description of our preliminary results.

The unknown spectrum is a sum

$$S(e^{j\omega}) = \sum_{m=-\infty}^{\infty} R_m e^{-jm\omega} \tag{20}$$

where R_m is known only for $m \in D$. In the MEM, the unspecified values of R_m are such as to maximize the entropy rate[28]

$$H = \int_{-\pi}^{\pi} \ln S(\omega)e^{j\omega}d\omega \tag{21}$$

of the underlying process $\underline{x}[n]$. This condition yields

$$\frac{\partial H}{\partial R_m} = 0 = \int_{-\pi}^{\pi} \frac{1}{S(e^{j\omega})} e^{-jm\omega}d\omega \qquad m \in D \tag{22}$$

from which it follows that

$$\frac{1}{S(e^{j\omega})} = \sum_{m\in D} c_m e^{-jm\omega} \tag{23}$$

The above suggests the following steepest descent method for maximizing H.

We denote by M the last integer in the data set D and we construct a positive definite set of M + 1 numbers R_m meeting the given data. We do so using a method based on the equivalence between reflection coefficients and positive definite sequences. We next construct an all-pole spectrum of order M whose inverse is the above sequence R_m. This spectrum is of the form (23) where all coefficients c_m are different from zero. Our objective is to find a sum as in (23) where $c_m = 0$ for $m \in \bar{D}$. To do so, we modify the unspecified values of R_m such that

$$\Delta R_m = \delta c_m \qquad m \in \bar{D} \tag{24}$$

and iterate the process. As we see from (22), this is the steepest descent solution to our problem. We should note that the underlying numerical operations are simple, involving only Levinson's algorithm.

Work in progress involves the convergence of the iteration under various conditions and a variety of numerical results.

5. REFERENCES

1. R.B. Blackman and J.W. Tukey, The Measurement of Power Spectra, Dover, New York, 1959.

2. G.M. Jenkins and D.G. Watts, Spectral Analysis and Its Applications, Holden-Day, San Francisco, 1968.

3. A. Papoulis, "Minimum Bias Windows for High Resolution Spectral Estimates," IEEE Transactions on Information Theory, Vol. IT-19, pp. 9-12, 1973.

4. A. Papoulis, "Apodization for Optimum Imaging of Smooth Objects," J. Opt. Soc. Am., Vol. 62, No. 12, pp. 1423-1429, December 1972.

5. A. Papoulis, "Estimation of the Average Density of a Non-uniform Poisson Process," IEEE Transactions on Communications, Vol. COM-22, No. 2, February 1974.

6. D. Slepian, H.J. Landau, and H.O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty Principle I and II," Bell Syst. Tech. J., Vol. 40, No. 1, pp. 43-84, 1961.

7. A. Papoulis and M. Bertran, Digital Filtering and Prolate Functions," IEEE Trans. Circuit Theory (Special issue on nonlinear circuits), Vol. CT-19, pp. 674-681, November 1972.

8. A. Papoulis, "A New Method of Image Restoration," Joint Services Technical Advisory Committee, Report No. 39, 1973-74.

9. A. Papoulis, "A New Algorithm in Spectral Analysis and Bandlimited Extrapolation," IEEE Transactions on Circuits and Systems, Vol. CAS-22, 1975.

10. A. Papoulis and C. Chamzas, "Improvement of Range Resolution by Spectral Extrapolation," Ultrasonic Imaging 1, pp. 121-135, 1979.

11. A. Papoulis and C. Chamzas, "Detection of Hidden Periodicities by Adaptive Extrapolation," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, No. 5, October 1979.

12. W. Xu and C. Chamzas, "On the Extrapolation of Band Limited Functions with Constraints," IEEE Trans. Acoustics and Signal Processing, October 1983.

13. P.S. Naidu and B. Paramasivaiah, "Estimation of Sinusoids from Incomplete Time Series," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, No. 3, June 1984.

14. E.T. Jaynes, "Prior Probabilities," IEEE Trans. Syst. Sci., Cybern., Vol. SSC-4, 1968.

15. T.J. Ulrych and T.N. Bishop, "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," Rev. Geophys. Space Phys., Vol. 13, February 1975.

16. J.A. Cadzow and R.L. Moses, "An Adaptive ARMA Spectral Estimator, Parts 1 and 2," Proceedings 1st Workshop on Spectral Estimation, McMarten University, Canada 1982.

17. J.E. Shore and R.W. Johnson, "Axiamatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," IEEE Trans. Inform. Theory, Vol. IT-26, pp. 26-37, January 1980.

18. Rodney W. Johnson, John E. Shore, "Multisignal Minimum-Cross-Entropy Spectrum Analysis with Weighted Initial Estimates," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, No. 3, June 1984.

19. E.T. Jaynes, "Prior Information and Ambiguity in Inverse Problems," AMS-SIAM Symposium on Inverse Problems, New York, New York, April 13, 1983.

20. B. deFinetti, "Prevision: Its Logical Laws, Its Subjective Sources," Translated from the French in H.E. Kyburg and H.I. Smokler, Studies in Subjective Probability, 2nd Edition, J. Wiley & Sons, Inc., New York, 1981.

21. S.W. Wilkins, J.N. Varghese, and M.S. Lehmann, "Statistical Geometry: A Self-Consistent Approach to the Crystallographic Inversion Problem Based on Information Theory," Acta Cryst. A39, pp. 47-60, 1983.

22. A. Papoulis, "Maximum Entropy and Spectral Estimation: A Review," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-29, 1981.

23. J. Proakis, Introduction to Digital Communication, McGraw-Hill, New York, 1970.

24. R. Kubo, "A Stochastic Theory of Line-Shape and Relaxation," Scottish Universities Summer School, D. ter Haar, ed., Plenum Press, New York. See also A. Papoulis, "Spectra of Stochastic FM Signals" in Proceedings of Transactions of 9th Prague Conference on Information Theory, 1982.

25. A. Papoulis, "Spectra of Stochastic FM Signals," in Proc. Trans. 9th Prague Conf. Inform. Theory, June 1982.

26. A. Papoulis, "Random Modulation: A Review," IEEE Transactions on Acoustics, speech, and Signal Processing, Vol. ASSP-31, No. 1, February 1983.

27. J.P. Burg, "Maximum Entropy Spectral Analysis," presented at the Int. Meeting Soc. Explor. Geophys., Orlando, Florida, 1967.

28. A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw Hill, New York, 2nd edition, 1984.

29. A. Papoulis, "Levinson's Algorithm, Wold's Decomposition, Spectral Estimation," SIAM Journal (to appear) 1985.

30. V.F. Pisarenko, "The Retrieval of Harmonics," Geophysical Journal of the Royal Astronomical Society, 1973.

31. G. Szego, "Orthogonal Polynomials," Amer. Math. Soc., Colloquium Publications, 1939.

32. D.C. Youla, "The FEE: A New Tunable High-Resolution Spectral Estimator," Polytechnic Tech. Note No. 3, Contract No. F30602-78-C-0048, RADC, 1978.

33. J.S. Lim, N.A. Malik, "Maximum Entropy Power Spectrum Estimation of Signals with Missing Correlation Points," ASSP-29, December 1981.

B. ROBUST METHODS IN PATTERN RECOGNITION, IMAGE PROCESSING, AND CLASSIFICATION AND ESTIMATION PROBLEMS

Professor L. Kurz

Unit IE4-2

1. OBJECTIVE(S)

In a paper which reviews twenty years of accomplishments in pattern recognition and image processing, Fu and Rosenfeld[1] state that in the decision-theoretic approach to image processing problems "we are still looking for effective and efficient feature extraction and selection techniques, particularly in nonparametric and small sample situations" [1, p. 365]. The latter class of problems is one of the major thrusts of this program. Though Rosenfeld and Fu refer to nonparametric methods, in view of recent developments it would be appropriate to replace the word "nonparametric" by "robust." In the engineering literature the nonparametric property was understood to mean constant false alarm rate (CFAR), usually disregarding the efficiency of the procedure. On the other hand, robust procedures stress efficiency and insensitivity to variations of the underlying noise distributions; they may or may not possess the CFAR property but in all cases a guaranteed upper bound on the false alarm rate is achievable. It should be noted that true robustness over a broad range of variations in environmental conditions is not possible without some form of adaptability which leads naturally to the need for robust estimation procedures which can continuously estimate the parameters required to maintain robustness of operation. Therefore, further work in the area of robust estimation is of prime importance. In particular, the interaction of robust classification or detection procedures with robust estimation techniques in adaptive processes and systems will be studied. The investigation will encompass three philosophically different classes of techniques: methods based on experimental design techniques, methods based on statistical partition and related tests, image reconstruction and related algorithms. As indicated by Fu and Rosenfeld,[1] these techniques have applications in communications (data compression, speech recognition), medicine (diagnosis, abnormality detection), and automation (robot vision). These techniques also have implications in such areas as underwater and earth sounding, radar maps, search radars, remote sensing, identification of human faces and fingerprints, reliability, machine part recognition, etc.

2. APPROACH

Unlike the approaches used by other researchers to some problems in pattern recognition, image processing and classification -- which concentrated in the past on spectral estimation, Bayesian decision-theoretic, linguistic and ad hoc procedures -- the classes of approaches proposed here are essentially statistical with robust properties. Even if the knowledge of underlying statistics of noise is limited, or changed in the processing or acquisition of data, the methodology outlined in this proposal yields satisfactory performance. Thus, the stress in the approach is on the robustness of the statistical methods -- good performance under varying or poorly defined noise conditions. In particular, the robustizing approaches based on the m-interval polynomial approximation (MIPA) concept also have implications in such areas as robust recursive estimators, robust spectral estimation and system identification and robust p.d.f. and c.d.f. estimation. Improvements in the MIPA approach lead to less partitions and improved approximation per interval. The latter philosophy leads to improvement in performance while decreasing processing time.

The stress is placed on the flexibility of designs and proper optimization in the presence of conflicting physical restrictions on the operation of systems being investigated.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

The problem of object detection based on the use of F- and t-statistics has been refined, and the results were submitted to a journal for pre-publication review. Trajectory detection procedures, based on symmetrical balanced incomplete block (SBIB) design masks to generate designs which produce the sharpest results for a preselected level of testing (confidence), were submitted and accepted for publication by the Journal of Computer Vision, Image Processing and Computer Graphics. The corrected galley proofs are available upon request.

The investigation into the area of robust array detection, initiated in reference 33, was extended to more general models. The m-interval polynomial approximation (MIPA) approach to robust detection, estimation, classification, pattern recognition and feature extraction was refined and extended to sequential operations of the processors. The performance of the MIPA methodology as compared to the min-max approach suggested by Huber was given for fixed and sequential modes of operation.

Extension of the algorithmic approach to the image reconstruction problem, initiated in reference 18, to two dimensions involving Walsh and other orthogonal functions in the two-dimensional reconstruction process, were studied. The pertinent algorithms were robustized using three principles: fixed structure robustizer, Mann-Whitney preprocessor and MIPA robustizer.

4. STATE OF THE ART AND PROGRESS DETAILS

This work is in part a natural continuation of some of the work supported previously by the JSEP (and builds upon former accomplishments resulting from that support), and in part consists of completely new material. In particular, material involving masking and related methods based on experimental design techniques, and involving some classes of recursive stabilizer algorithms for image reconstruction, are an outgrowth of the previous contract. Material relating to nonlinear partition classifiers follows from former material. On the other hand, other material is completely new, and is only in general terms related to the previous approaches to the problems. In this category we should mention the concept of the m-interval polynomial approximation (MIPA) approach to classification and estimation problems, the statistical approach to object detection by efficient utilization of F- and t-statistics simultaneously, and the use of F-statistics in gradient-type algorithms such as in problems involving the reconstruction of poorly defined image edges and cracks.

The material is grouped into four clusters of problems: 1. Methods based on experimental design techniques.[2-4] Initial efforts in this area were reported in references 5-8. These problems involve the design of appropriate statistical masks which base their sensitivity on the appropriate F-statistic. The problems considered are object, trajectory and edge detection, and image enhancement. 2. Methods based on partition and related statistical tests. A summary of procedures based on partition tests up to the year 1977 is given in reference 9. Extension of these techniques to other forms of partition tests,[10 15] with particular stress on two-dimensional,[11] two-sample (learning with a teacher or supervised learning) and nonlinear statistics, will be considered. A new approach to generation of robust partition classifiers or detectors based on the concept of m-interval polynomial approximation (MIPA) method, introduced by the senior investigator in connection with some classes of estimation[16] and identification[17] problems, represents a particularly interesting and fruitful approach to classification problems in poorly defined or variable noise environment. 3. Image reconstruction algorithms. The proposed algorithms belong to two classes -- recursive stabilizer algorithms for image reconstruction[18] and maximum gradient algorithms based on F-statistics (finding best edge approximations to cracks). 4. Robust recursive estimators. Though the need of robust recursive estimators has been demonstrated in the past,[18-22] a new class of such estimators is introduced here because these do not require inconvenient rank preprocessing and base their robustness and rapid convergence on the MIPA concept. Additional stress is placed on sequential procedures which are particularly useful in poorly defined and changing noise environments.

A. Methods Based on Experimental Design Techniques

(1) Object Detection and Experimental Designs

When ANOVA is applied to image data, a set of F-statistics is generated.[6-8] These statistics are related to confidence regions and can be used to determine if there exist significant row, column effects or interaction between them. Additional information can be incorporated to determine, with a certain degree of confidence, if there exists significant contrast between a pair of rows or columns. Thus, ANOVA transforms an image into a set of F-statistics. The ANOVA statistics are in general insufficient to describe fully the pattern of an image because it confirms only three possible edges: vertical, horizontal and diagonal. Nothing can be concluded if the pattern is described by edge elements other than the three mentioned above. This has been one of the major handicaps in extending ANOVA techniques to scene analysis. However, there exists a pattern which can be described by its ANOVA statistics. Consider the pattern (array) of Figure 1. It is obvious that this pattern has two parts and each part is homogeneous. This leads to the definition of a standard pattern. Let S be an $m \times m$ array, B be its subarray which includes columns 1 to k, $1 \leq k \leq m$, and T be the remaining array. Also, let α denote the confidence coefficient. The column and row F-statistics of S, B and T are denoted by FS_c, FS_r, FB_c, FB_r, FT_c and FT_r, respectively. The confidence interval of a contrast between columns k and k + 1 is[23]

$$(c_k, c_{k+1}) = \bar{X}_{\cdot k} - \bar{X}_{\cdot (k+1)} \pm t_{v,1-\alpha/2}(2MS_e/m)^{\frac{1}{2}}.$$

111111000
111111000
111111000
111111000
111111000

Fig. 1

Definition A.1: S is a standard pattern if the F-statistics of ANOVA and t-values of contrasts satisfy: (a) S is heterogeneous, i.e., $FS_c \geq V_\alpha$ and $FS_r < F_\alpha$; (b) B is homogeneous, i.e., $FB_c < F_\alpha$ and $FB_r < F_\alpha$; (c) T is homogeneous, i.e., $FT_c < F_\alpha$ and $FT_r < F_\alpha$; (d) contrast exists between columns k and k + 1, i.e., (c_k, c_{k+1}) does not cover zero.

For example, the statistics of the array of Figure 2 are FS_c = 105.366, FS_r = .935, FB_c = .727, FB_r = .848, FT_c = 1.369, FT_r = .867. The threshold values corresponding to 95% confidence level of S(9×16), B(9×5) and T(9×11) are TS_c = 1.751, TS_r = 2.016, TB_c = 2.687, TB_r = 2.265, TT_c = 1.951 and TT_r = 2.057. The contrast confidence interval is (c_5, c_6) = - 1.934 and -2.288. By comparing the calculated statistics and the threshold values, it is obvious that all four conditions are satisfied. This is a standard pattern of 9 rows and 16 columns with the border occurring between columns 5 and 6 with a 95% confidence level. A standard pattern is a function of four variables: the row size (m) and the column size (n) of the array, the border column (k) in the B subarray and the confidence coefficient (α). A given set of values (m,n,k,α) describes a group of standard patterns for which the confidence levels are equal or higher than α. This is true because α defines a limit and F-values are monoincreasing in nature. All patterns in such a set produce the same visual effects, the only difference being the relative intensity of two subarrays.

In order to apply ANOVA to an array of arbitrary patterns, one would transform the original array into a standard pattern. However, the transformation rules must be derived in a random fashion. Randomization process destroys any a priori knowledge of a given model and, therefore, provides observations free of bias. Let W be such a set of transformations which is one-to-one and onto; namely, $W:S(m \times n) \rightarrow R(p,q)$, where mn = pq. One can apply W to any array of size m×n and then analyze the transformed array by ANOVA to see if the original array is transformable to a standard pattern. We can now define the visual equivalence of two arrays.

Definition A.2: Let arrays A and B of the same size m×n and W be a set of randomly generated transformation rules such that $W:A(m \times n) \rightarrow A'(p \times q)$; $W:B(m \times n) \rightarrow B'(p \times q)$. A and B are visually equivalent, with a confidence coefficient α, if both A' and B' are standard patterns.

Note that visual equivalence is based on the statistics of the transformed array, not the original one. Since all standard patterns characterized by the same set of parameters (p,q,k,α) give the same visual effects and the transformation process is one-to-one, one expects the original arrays to exhibit visual resemblance.

```
4444466666676666
4444467667666666
4445466676676666
5444466766666666
4444466766666667
4444466766766676
4444466766666666
4444466666666666
4444466666666666
```

Fig. 2

Consider the problem of object detection in the framework of supervised learning. The task is to design a classifier with the knowledge obtained from a given set of samples called teachers or templates. Each template is composed of a model object, the target and its surroundings, the background. The background portion is needed to define the boundary and to increase the relative intensity of the target. Figure 3 shows a proposed recursive learning scheme utilizing ANOVA, randomized transformation rules and the standard pattern. The criteria of terminating the recursive process are the F- and t-statistics associated with a predetermined confidence level. As shown in the figure, the learning process starts by "coloring" the three components in a template array: target, background pixels and those not to be included in the transformation. The next step is to propose the size of a standard pattern by finding the common denominators of the total number of target pixels and background pixels. Let d_i be such a common denominator or the product of a group of nonrepeating common denominators. Also, let n_t and n_b be the total number of target and background pixels, respectively. Then $p = d_i$, $q = (n_t + n_b)/d_i$ and $k = n_b/d_i$. Even though any d_i can be justified as the final choice of p, it is advised to select d_i to minimize the difference between p and q. In step 3, every target pixel is assigned to a position in the target subarray of the standard pattern. The rule to follow is that the assignments should be one-to-one and onto. To ensure randomness, a random number generator is employed to relate the "from" and "to" pixels. All successful mappings are recorded as transformation rules. The assignments of background pixels to the background subarray are done similarly. In step 4, the transformation rules are applied to the original array and the resultant array is analyzed by ANOVA. If the F- and t-statistics do not qualify to be a standard pattern, it is clear that the

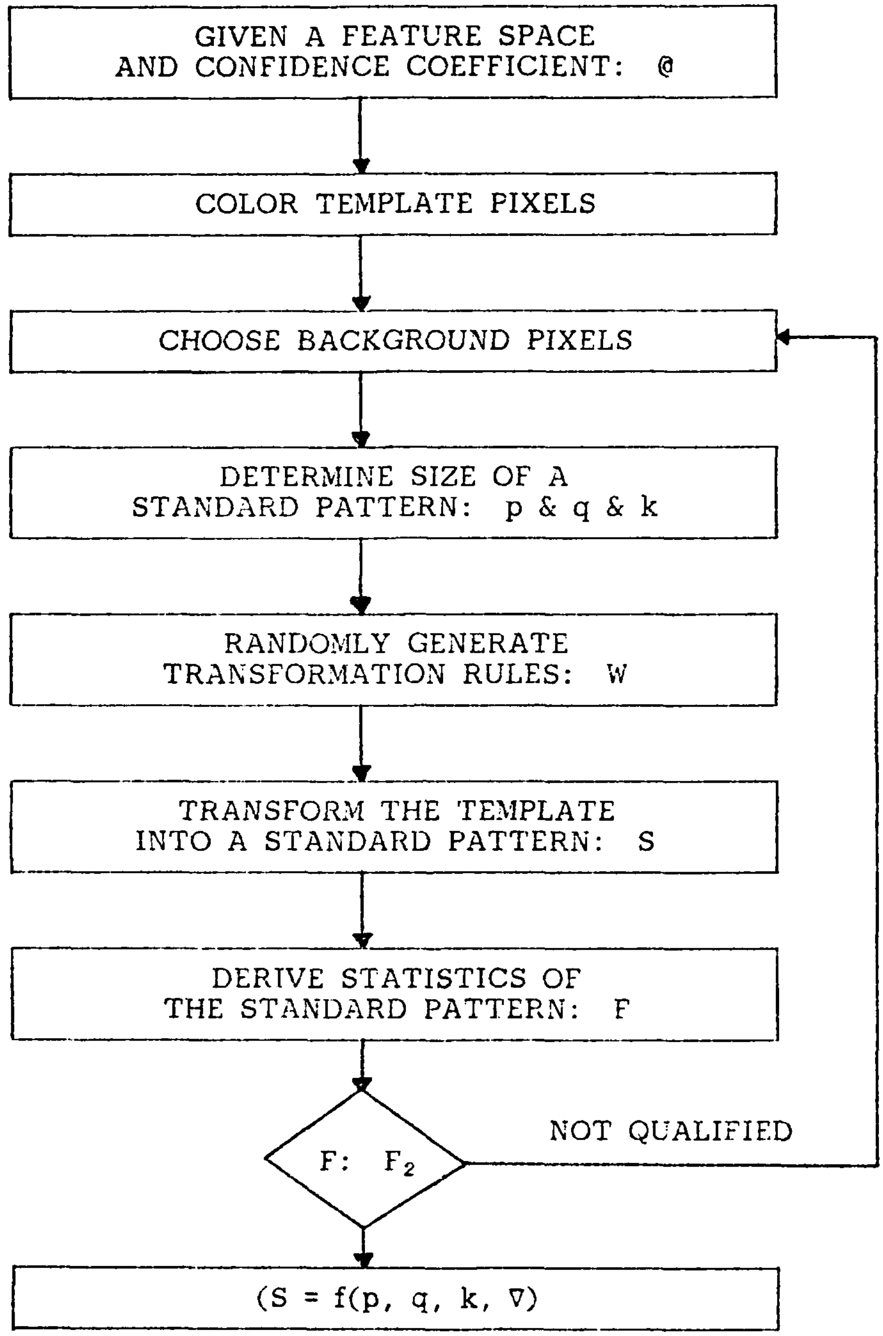


Fig. 3. Supervised learning scheme.

** THE TEACHER FILE ::	** THE TRANSFORMED ARRAY **
1111111111	1111111353
1111112111	1111111333
1133354311	1111111333
1113343311	1112111333
1133333311	1111121334
1133353311	1111111435
1123331111	1112111333
1113533111	1111111333
1111111111	1111111353
1111111211	1111111333

** THE TRANSFORMATION RULES **

```
( 1, 1)-( 5, 7) ( 1, 2)-( 5, 1) ( 1, 3)-( 5, 4) ( 1, 4)-( 1, 1) ( 1, 5)-( 1, 3)
( 1, 6)-( 3, 7) ( 1, 7)-( 9, 7) ( 1, 6)-( 2, 7) ( 1, 9)-(10, 5) ( 1,10)-( 7, 4)
( 2, 1)-( 4, 7) ( 2, 2)-(10, 2) ( 2, 9)-( 1, 5) ( 2, 4)-( 5, 3) ( 2, 5)-( 2, 1)
( 2, 6)-(10, 3) ( 2, 7)-( 7, 4) ( 2, 8)-( 9, 4) ( 2, 9)-( 2, 8) ( 2,10)-( 8, 7)
( 3, 1)-( 1, 7) ( 3, 2)-( 2, 3) ( 3, 6)-( 8,10) ( 3, 4)-( 2, 9) ( 3, 5)-( 4, 5)
( 3, 6)-( 1, 9) ( 3, 7)-( 5,10) ( 3, 8)-( 8, 8) ( 3, 9)-( 2, 5) ( 3,10)-( 7, 8)
( 4, 1)-( 4, 5) ( 4, 2)-( 8, 5) ( 4, 2)-( 8, 8) ( 4, 4)-(10, 9) ( 4, 5)-( 5, 6)
( 4, 6)-( 6, 8) ( 4, 7)-(10, 9) ( 4, 8)-(10,10) ( 4, 9)-( 3, 9) ( 4,10)-(10, 4)
( 5, 1)-(10, 7) ( 5, 2)-( 8, 1) ( 5, 3)-( 9,10) ( 5, 4)-( 1, 8) ( 9, 5)-( 3, 8)
( 5, 6)-( 7, 9) ( 5, 7)-( 2,10) ( 5, 8)-( 1,10) ( 5, 9)-( 8, 2) ( 5,10)-( 8, 2)
( 6, 1)-( 3, 1) ( 6, 2)-( 3, 2) ( 8, 9)-( 7,10) ( 6, 4)-( 6, 9) ( 6, 5)-( 9, 1)
( 6, 6)-( 9, 9) ( 6, 7)-( 2, 8) ( 8, 8)-( 4, 8) ( 6, 9)-( 7, 3) ( 6,10)-( 8, 4)
( 7, 1)-( 7, 2) ( 7, 2)-( 9, 3) ( 7, 3)-( 5, 8) ( 7, 4)-( 4, 9) ( 7, 5)-( 5, 8)
( 7, 6)-( 3,10) ( 7, 7)-( 3, 5) ( 7, 8)-( 8, 3) ( 7, 9)-( 8, 4) ( 7,10)-( 6, 5)
( 8, 1)-( 8, 2) ( 8, 2)-( 1, 4) ( 8, 3)-( 6, 5) ( 8, 4)-( 8, 9) ( 8, 5)-( 6, 3)
( 8, 6)-( 7, 8) ( 8, 7)-( 3, 9) ( 8, 8)-( 5, 5) ( 8, 9)-( 8, 4) ( 8,10)-( 9, 3)
( 9, 1)-( 1, 8) ( 9, 2)-( 9, 2) ( 9, 3)-( 9, 8) ( 9, 4)-( 8, 4) ( 9, 5)-( 4, 8)
( 9, 6)-( 9, 1) ( 9, 7)-( 5, 2) ( 9, 8)-( 2, 1) ( 9, 9)-( 2, 2) ( 9,10)-( 3, 5)
(10, 1)-( 8, 7) (10, 2)-(10, 1) (10, 3)-(10, 4) (10, 4)-( 4, 6) (10, 5)-( 7, 5)
(10, 6)-( 7, 6) (10, 7)-( 4, 2) (10, 8)-( 4, 4) (10, 9)-( 1, 3) (10,10)-( 4, 3)
```

: THE CONTRAST BETWEEN COLUMNS :

COLUMN 1	COLUMN 2	CONTRAST
1	2	0
2	3	0
3	4	0
4	5	0
5	6	0
6	7	0
7	8	1
8	9	1
9	10	9

	S_c	S_r	B_c	B_r	T_c	T_r
OBSERVED:	73.553	0.776	1.519	0.818	0.558	1.027
THRESHOLD:	1.999	1.999	2.040	2.040	2.460	2.460

Fig. 4

chosen background does not provide enough visual contrast to make the target stand out. A different template or a different set of background pixels should be tried. The process is then repeated from the first step. If the first template is found to be transformable to a standard pattern of confidence level α, the learning process is completed. Figure 4 shows a template, the transformation rules, the standard pattern and all the statistics associated with them for the confidence level 95%.

Once a template is represented by a set of statistics, the task of object detection becomes a straightforward problem in hypothesis testing based on ANOVA. As shown in Figure 5, an image array is transformed and the resultant array is analyzed by ANOVA. If the F- and t-statistics qualify as belonging to a standard pattern of confidence coefficient α, the image data is said to be visually equivalent to the template.

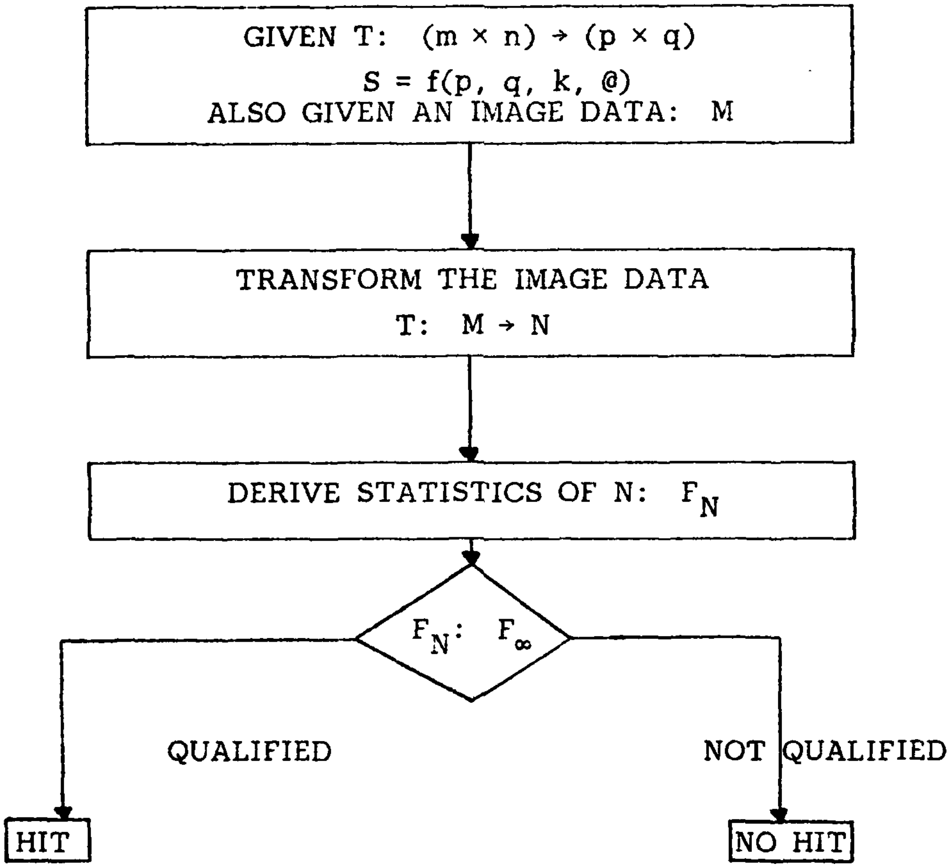


Fig. 5. Object detection scheme.

It is proposed to continue research in this area stressing the following items: Study the convergence properties of the proposed object detection scheme, extend the procedure to multiframe object detection, and develop a coordinate-free transformation rules generator in the learning scheme. Basic results in this area were submitted for publication.[45]

(2) Trajectory Detection and Experimental Designs

A trajectory is identified by a continuous and narrow region of high intensities which is surrounded by a region of low intensities on each side. When an image containing such a trajectory is processed by a localized operator (window or mask), the grey level distribution within the mask behaves like either a pulse function or a step function, depending on whether the mask is placed perpendicularly or in parallel to the trajectory. The traditional square-shaped edge detector is ineffective in this application regardless of the relative position of the mask to the trajectory. A small edge detector may not contain enough variations due to trajectory elements to cause significant block effects. On the other hand, a large edge detector may cover too wide an area so that adjacent trajectory segments are included in the same mask. What is needed is an experimental design in the shape of a stretched rectangle. When it is placed perpendicularly, it allows for the inclusion of three component parts of a pulse function. When it is placed in parallel, it allows for the inclusion of substantial elements from only one trajectory segment to form a step function.

A trajectory is further characterized by two inherent properties: the block nature and strong contrast of trajectory elements. Firstly, those elements containing a trajectory tend to be adjacent to one another. Their grey level or texture tend to be similar when compared with the surrounding elements on both sides. Secondly, the grey level distribution changes drastically around both edges of the trajectory. So only those block effects which are associated with strong contrast can be considered as indicative of a "true" trajectory. Based on these properties, an image data can be considered as a three-dimensional array with the row and column numbers describing the planar position, and the grey level distribution is used as the third dimension to describe the relative contrast. In terms of experimental designs, the analysis of an image containing a trajectory calls for a model which can determine the presence of row, column and grey level effects independently.

A SBIB design has the shape of a rectangle due to its definition that the number of blocks is greater than the number of units in each block. Its analysis determines if there exist significant block or grey level effects independently. It also derives the adjusted block means to be used as the unbiased estimators of the true block means.[26] In its application to the trajectory detection problem, the hypothesis that a SBIB mask contains a trajectory is accepted if the following three conditions are all satisfied:

(a) There exist significant block effects.
(b) There exist significant grey level effects.
(c) The adjusted block means behave like a pulse function.

The presence of block effects indicates that there exist edges. When the block effects are qualified by grey level effects, only those edges of strong contrast are preserved. The study of the adjusted block mean indicates that it differentiates a trajectory from an edge.

To satisfy the randomization criterion, a symmetrical balanced incomplete block (SBIB) design is selected from a set of existing designs randomly and then its grey level designations are further randomized by permutations of the row and column assignments. The first two conditions can be easily implemented by comparing the F-statistics of a SBIB design with a threshold value corresponding to certain confidence levels. The shape of the grey level distribution can be determined by thresholding the adjusted block means by their overall means.

(3) Edge Detection Using Radial F-Statistics

Detecting edges separating grey levels is an important aspect of image processing. Here a procedure for detecting edges based on radial F-statistics is proposed. Though the procedure is uniformly most powerful (UMP) for gaussian noise background,[25] it remains robust (insensitive to changes in noise distribution) in many situations.[26] If conditions of reference 26 are not satisfied, robustness of the procedure may be achieved by robustizing algorithms of reference 7. The actual detection method is based on using F-statistics along radial vectors from the inside of the object enclosed by the edge. It can be shown that the radial F-statistic varies linearly between two constant values in passing between two grey levels. The midpoint of the linear portion represents a true edge point. The F-statistic is of the form

$$F_{i;m} = \frac{V_i^2}{V_t^2} = \left[\sum_{i=1}^{m} (S_i - \Sigma S_i/m)^2/m\right] \div \left[\sum_{i=1}^{m} (S_{ti} - \Sigma S_{ti}/m)/m\right]$$

where V_i^2 = sample variance of the extracted sample; V_t^2 = sample variance of the training sample, and the notation of reference 8 has been used. A paper on this subject has been submitted for publication.[46]

(4) Image Enhancement

The theory of Latin squares which was applied to edge detection by Kadar and Kurz[6] also can be applied successfully to image enhancement. Presentation of the methodology is meaningless without demonstration on actual images. Since the proper images are available only on film, the description of this method will be postponed to the oral presentation.

B. Methods Based on Partition and Related Statistical Tests

(1) Classifiers or Detectors Based on Occupancy Vectors

In this section, the theory of m-interval detectors[9] is extended to two-sample detectors based on occupancy vectors. The objective is to find simple, robust detectors for signals in noise with unknown or poorly defined statistics. The Taylor series expansion of the likelihood ratio based on the occupancy vector distributions is used to derive linear and quadratic two-sample tests which are locally most powerful. The first term of the expansion leads to a linear test which is locally more powerful if the relationship of the distributions under the hypothesis (H_0) and the alternative (H_1) is known, i.e., shift of the mean, stochastic ordering, etc. If the relationship is now known, the second term of the expansion leads to a quadratic test which is sensitive to any differences between the received and the reference occupancy vectors. Detectors based on occupancy vectors require specification of m constants $\{a_1, a_2, \ldots, a_m\}$ to partition the space of the distribution $F(x)$ under H_0. Most frequently these constants are chosen to be quantiles $a_i = F^{-1}(1/m)$ which can be easily estimated by methods introduced by the senior investigator.[11,20] Choosing $\{a_i\}$ not coinciding with quantiles may be useful in some applications, especially if the nature of the relationship between H_0 in H_1 is unknown. If $\{x_i\}$ represents the received data (sample) and $\{y_k\}$ the reference or normalization sample, the components of the occupancy vectors for disjoint intervals $A_i = (a_{i-1}, a_i]$ are defined by

$$h_i = \frac{1}{L} \sum_{i=1}^{L} I_{A_i}(x_i)$$

for the received sample and

$$\eta_i = \frac{1}{K} \sum_{i=1}^{K} I_{A_i}(x_k)$$

for the normalization sample. The occupancy vectors for the received (test) and normalization (reference) samples are then $\underline{h}=(h_1, h_2, \ldots, h_m)$ and $\underline{\eta} = (\eta_1, \eta_2, \ldots, \eta_m)$, respectively. It can be shown that the locally most powerful two-sample linear m-interval detector (TLMID) is of the form

$$T = \sum_{i=1}^{m} b_i(h_i - \eta_i) \qquad \text{(B.1)}$$

where the weights (scores) $\{b_i\}$ are selected subject to some optimization criterion. It is proposed to use the generalized signal-to-noise ratio (GSNR). The GSNR for a test T is defined as $GSNR(T) = \{E[T/H_1] - E[T/H_0]\}^2/var(T/H_0)$ which, for small signal case ($\theta \to 0$), reduces to

$$\lim_{\theta \to 0} GSNR(T) = \left[\frac{d}{d\theta} E[T/H_1]|_{\theta \to 0}\right]^2 / var(T/H_0)$$

The small S/N GSNR is known as the efficacy. The TLMID gains its robustness from the relative insensitivity to changes in $F(x)$. Crucial to the determination of $\{b_i\}$ is the relationship of $F(x)$ to the distribution under the alternative, $G(x)$, i.e., $G(x) = F(x-\theta)$, $\theta \geq 0$ (shift-of-mean); $G(x) = F(x/(1+\theta), \theta \geq 0)$ (change-of-scale); $G(x) = F^{1+\theta}(x)$, $\theta \geq 0$ (Lehmann alternative), etc. It can be shown, after considerable mathematical manipulations, that for the TLMID the optimum choice of b_i is $b_i = p_i(0)$ where

$$p_i(0) = \frac{d}{d\theta}[F_\theta(a_i) - F_\theta(a_{i-1})]|_{\theta=0}$$

and the corresponding optimum GSNR(T) is

$$\mathrm{GSNR(T)} = (LK/L + K)\ m \sum_{i=1}^{m} [p_i(0)]^2 \qquad (B.2)$$

In a straightforward manner one can show that the statistic of (B.1) is asymptotically gaussian. Thus, one can compare TLMID with other detectors based on the asymptotic relative efficiency (ARE). If the noise is gaussian, ARE of TLMID with respect to the optimum parametric detector (OPD) is

$$m \sum_{i=1}^{m} [f(a_{i-1}) - f(a_i)]^2$$

which is the same as for the one-sample m-interval detector.[9] Thus, for m = 6 TLMID reaches over 90% efficiency of the OPD if the noise is gaussian and remains efficient even if burst noise occurs, or the TLMID is robust. In spectral estimation problems the interest is concentrated on the change-of-scale alternatives.[27] Particularly, $g(x) = 1/2^B\ \Gamma(B)x^{B-1}e^{x/2}u(x)$ which is the gamma distribution. Following Woinsky,[27] the efficacy of the OPD is $\varepsilon(t_2) = (LK/L+K)B$, which, compared to the efficacy of the TLMID, yields

$$ARE_{T/t_2} = m/B \sum_{i=1}^{m} [a_{i-1}\ f(a_{i-1}) - a_i f(a_i)]^2 \quad ;$$

for $B \geq 1$ and $m \geq 5$ ARE rapidly approaches one. Thus, the robustness, simplicity of implementation, and efficiency make the TLMID particularly attractive in spectral estimation problems.[27]

The two-sample LMID provides robustness in two forms: by the two-sample detection and by partitioning. The two samples provide robustness against changes in the distribution of the hypothesis of a type similar to the alternative. For instance, in the shift alternative case, the two-sample test maintains performance even if the mean of the background drifts away from zero, where a one-sample test would fail. The second type of robustness is common to partition tests and is due to the unrefined form of partitioning of the space.

In problems where the relationship between F(x) and G(x) is unknown, i.e., poorly defined dispersive media, the two-sample LMID is replaced by a two-sample quadratic m-interval detector (TQMID) which corresponds to the second term of the Taylor's series expansion. A practical form of the TQMID is

$$T_q = \sum_{i-1}^{m} c_i(h_i - \eta_i)^2 \tag{B.3}$$

If $c_i = 1/p_i$, where p_i is the probability of x_i falling into a given occupancy cell, T_q of (B.3) reduces to the well-known goodness-of-fit test. To find the best weight vector $\{c_i\}$, the approach based on GSNR is not useful. The maximization of the GSNR requires knowledge of the form of the alternative and in the end yields a test sensitive to only that form of alternative. The potential applications of the quadratic test is to detection problems where this knowledge is not available. The weights $\{c_i\}$ are chosen to minimize the variance of T_q of (B.3) under the hypothesis. This approach is related to the one suggested in reference 11. It can be shown that for quantile partitions the optimum $c_i = -(1+2a)/2kb$ where $a = (1/m)(1-1/m)$, $b = (2/m^4 - 3/m^3 + 1/m)$ and

$$k = \sum_{i=1}^{m} c_i \ .$$

Choosing $c_i = 1/m$ makes the TQMID equivalent to a chi-square test with quantile partitions. For partitions other than quantile partitions, as would result from distributions other than the design hypothesis, the optimum weights are not equal. In the latter case, a mathematical relationship between the partition probabilities to the best weights can be used to adapt the detector to provide best performance. For example, a TQMID using fixed partitions in a slowly changing environment might use a periodic learning sample to update the weights to maintain minimum variance operation. This area requires additional work to find a general class of adaptive detectors.

By changing from invariable to bivariate (two-dimensional) statistics, one doubles the effective signal-to-noise ratio.[11] This is of particular importance in processing severely corrupted images. Therefore, the extension of the theory presented in this section to bivariate statistics is important. The extension presents some mathematical difficulties, but the senior investigator is confident that these difficulties can be overcome.

(2) The m-Interval Polynomial Approximation (MIPA) Method

Consider a function $g(x)$ which may represent a p.d.f., a c.d.f., an influence or scoring function in a robust detector, a discriminant function in pattern recognition, etc. The problem posed is:

Given a set of observables (samples) from a population corresponding to $g(x)$, find a robust, adaptive and efficient estimator, $\hat{g}(x)$, of $g(x)$. Consider $\hat{g}(x)$ of the form

$$\hat{g}(x) = \sum_{j=1}^{m} (c_{2j}\, x^2 + c_{ij}x + c_{0j})\, I_{A_j}(x) \tag{B.4}$$

where $I_{A_j}(\cdot)$ is the indicator function of the set $A_j = (a_{j-1}, a_j]$ and a_j are found from

$$\int_{a_{i-1}}^{a_j} g(x)dx = \lambda_j, \qquad \sum_{j=1}^{m} \lambda_j = \int_{-\infty}^{\infty} g(x)dx = \text{constant}, \quad a_0 = -\infty, \quad a_m = +\infty$$

such that some convenient measure of error is minimized, i.e., the integral square error (ISE)

$$I(c) = \int_{-\infty}^{\infty} [\sum_{j=1}^{m} (c_{2j}\, x^2 + c_{ij}x + c_{0j})\, I_{A_j}(x) - g(x)]^2\, dx \qquad \text{(B.5)}$$

is minimized with respect to $\underline{C}_j = (c_{2j}, c_{ij}, c_{0j})^t$. It can be shown that the optimum constants (scores) can be found from

$$\underline{c}_j = \underline{u}_j^{-1} \underline{\eta}_j$$

where

$$\underline{u}_j = \begin{bmatrix} u_{4j} & u_{3j} & u_{2j} \\ u_{3j} & u_{2j} & u_{1j} \\ u_{2j} & u_{1j} & u_{0j} \end{bmatrix} \qquad u_{kj} = \frac{1}{k+1}\,(a_j^{k+1} - a_{j-1}^{k+1});\ \frac{1}{k+1} \quad k = 0,1,2,3,4$$

$$\underline{\eta}_j = [\eta_{2j}, \eta_{1j}, \eta_{0j}]^t \qquad \eta_{ij} = \int_{a_{j-1}}^{a_j} x^i\, g(x)dx, \qquad i = 0,1,2$$

a_j's can be recursively estimated as in references 11 and 20 and η_{ij} can be estimated from

$$\omega_2^j(\ell+1) = \omega_2^j(\ell) + \frac{1}{\ell}\,[[x_\ell]_j^2 - \omega_2^j(\ell)]$$

$$\omega_1^j(\ell+1) = \omega_1^j(\ell) + \frac{1}{\ell}\,[[x_\ell]_j^1 - \omega_1^j(\ell)]$$

$$\omega_0^j(\ell+1) = \omega_0^j(\ell) + \frac{1}{\ell}\,[u(a_j - x_\ell) - u(a_{j-1} - x_\ell) - \omega_0^j(\ell)]$$

where $[x]_j = \begin{cases} x & \text{if } x \in A_j \\ 0 & \text{otherwise} \end{cases}$ and $u(\cdot)$ is the unit step function

All the estimators converge in m.s. sense and with probability 1. In an actual system all the parameters can be estimated simultaneously using a microprocessor operating in a parallel mode. Thus the procedure is continuously adaptive and can be used in a real-time system. It should be noted that the MIPA process outlined above can be modified by using other than quadratic polynomials in each interval or using different order polynomials in some intervals. The efficacy, robustness, and flexibility of this approach have already been demonstrated in other applications.[16,17] An alternate approach to this problem has been initiated. The new approach uses a smaller number of intervals but a better approximation per interval. The resulting modified techniques permit faster processing of data (smaller sample size needed for the same accuracy as in MIPA).

C. Image Reconstruction Algorithms

In a recent paper[18] the senior investigator introduced recursive algorithms which are useful in image restoration. Unlike methods suggested in the past, which result in ill-posed mathematical problems, the new algorithms converge in the mean square sense and with probability 1. In addition, the algorithms can be easily robustized (see Eq. (14) of reference 18). The immediate goal of future research in this area is extension of the algorithms to two dimensions involving Walsh functions in the two-dimensional reconstruction process. In addition, recent research by the principal investigators suggests that one can obtain a method of robustizing algorithms that is simpler than and as effective as the method described in reference 18. Thus, in extending the procedure to two dimensions the new robustizing process will be included. The methods will be refined to perform more efficiently in specific applications.

A parallel effort will be directed toward developing reconstruction algorithms based on the F-statistic. Consider the problem of edge reconstruction,[28,29] where most of the object was extracted by any suitable method such as the one presented in reference 8. Part of the edge which separates the object from the background is badly corrupted and not recognizable using standard masking techniques. One can think of the corruption as a localized burst noise. The radial F-statistic of Section A-3 can be modified to extract the desired edge. Let the image be distorted locally with the rest of the image noise-free. This renders the application of the standard F-statistic useless because the image elements are of zero variance and known mean. However, if a low-amplitude perturbation pattern, called dither, is added to the picture, the procedure could be applied taking the advantage of the robustness of the F-statistic. Consider a potential edge point S_n. The size of the mask centered on S_n is chosen to preserve the robustness of the statistic and to be able to follow the rate of change in the direction of edges to be classified. Starting with S_n, quantized directional vectors are chosen, and for each quantized directional vector, the gradient is computed at different points. Thus, the mask (image submatrix) is transformed into a differential submatrix of gradients falling below a preset threshold. The steps are as follows: gradients are computed for a number of points in an image containing no object but only background noise; based on the computed values, a threshold is chosen and this threshold is used to generate a mask. This mask is applied to the differential submatrix, and the point S_n is classified as belonging to an edge or not. If S_n is an edge point, the next point S_{n+1} is detected in the direction containing the largest gradient vector and having a location close to the most likely edge direction. Once the point S_{n+1} is detected, the point lying on the periphery of the submatrix is chosen as the next edge element to which the procedure is applied. This is repeated until a closed curve is obtained. If a closed curve does not appear, the procedure is repeated in the opposite direction and interpolation is used to extrapolate the edge. The new application of masks and differential operators can be used not only to reconstruct edges of a partially degraded image but in following the path of a moving object in a noisy environment and in related problems. Simultaneous use of object detection and gradient masks represents another fruitful area for further research. Comparing the relative operational advantages and disadvantages of the two reconstruction philosophies presented in this section will be considered.

D. Robust Recursive Estimators

In this section, a set of computationally attractive Robbins-Monro stochastic approximation (RMSA) procedures, based on the MIPA suggested in the previous section, are introduced. The estimators are asymptotically consistent, efficient, and robust. Extensions to other recursive algorithms are proposed. As a first task, it will be shown that there exists a RMSA equivalent to Huber's M-estimate.[30] Let $x_1, x_2, \ldots, x_n$ be i.i.d.r.v. and from a p.d.f. $f(x-\theta)$. Huber tried to estimate θ by either finding a minimum of

$$\sum_{i=1}^{n} \zeta(x_i - T_n)$$

or by implicitly solving

$$\sum_{i=1}^{n} \Psi(x_i - T_n) = 0 \qquad \text{(D.1)}$$

where ζ is an arbitrary well-behaved function, $\psi(x-\theta) = \frac{\partial}{\partial\theta} \zeta(x-\theta)$.

Solutions to Eq. (D.1) yield M-estimates. Under certain regularity conditions, T_n (an estimator of θ) is consistent and $n^{\frac{1}{2}}(T_n - \theta)$ converges in law to the gaussian distribution with mean zero and variance

$$V(\phi,f) = \frac{\int \psi^2 \, fdx}{(\int \psi^1 \, fdx)^2}$$

For n sufficiently large x,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \psi(x_i - T_n) = E_\theta[\psi(x-T)]$$

Smoothness of $f(x-\theta)$ ensures the existence of $\partial/\partial T$ and $\partial^2/\partial T^2$ of $f(x-T)$ and their absolute integrability. Let $\alpha(T) = E_\theta[\psi(x-t)]$, then

$$\alpha(T) = 0 \qquad T = \theta$$

$$\neq \text{ otherwise}$$

Thus, asymptotically the M-estimator is equivalent to finding the root of the regression function $\alpha(T)$ by using RMSA yielding

$$T_{n+1} = T_n + \frac{1}{n\alpha^1(T_n)} [\psi(x_n - T_n)] \tag{D.2}$$

where

$$\alpha^1(T_n) = E_\theta \frac{\partial\psi(x-T)}{\partial T} \; T = T_n \quad .$$

Following references 31 and 32, $T_n \to \theta$ in m.s. sense and w.p.l. and $n^{\frac{1}{2}}(T_n - \theta)$ is asymptotically normal of mean zero and variance $V(\psi,f) = \int \psi^2 \, fdx/[\int \psi^1 \, fdx]^2$, which is the same as Huber's result. If $\psi = f^1/f$, T_n is the maximum likelihood estimator (MLE) of θ and has the minimum variance

$$V(-f^1/f,f) = 1/\int(f^1/f)^2 \, fdx = \frac{1}{I_f}$$

where I_f is the Fisher Information of f. In robust estimation theory, f and f^1/f are not known. To circumvent this difficulty, Huber assumes that $f(x-\theta)$ belongs to

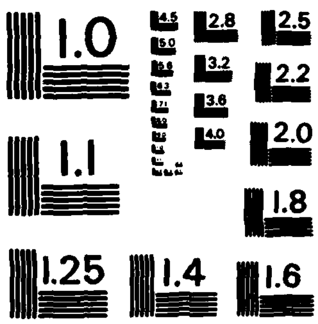
$$\Phi_s = \{f | f = (1-\varepsilon)g + \varepsilon h, \qquad h \text{ symmetric}\} \quad .$$

He obtains a min-max estimator by finding ψ which minimizes $\sup\{V(\psi,f):f \in \Phi_s\}$. It should be noted that without symmetry of h, the estimator is biased. Though the theory of min-max estimators is elegant and consistent, it is of limited practical use. Essentially, ambiguity about the knowledge of f has been carried over to the estimation problem of ε and h. Actually, without the knowledge of h, the crucial parameter ε cannot be estimated. It can be shown that wrong assumptions about the h lead to poor estimation of the mixing parameter ε, which is critical in yielding efficient M-estimators. In addition, since this is a nonlinear regression problem, the convergence rate is very slow and accurate estimates are almost never reached for finite sample sizes.

To bypass the difficulties associated with Huber's estimators, the theory of MIPA robust estimators is proposed by this investigator. In the subsequent material all assumptions are relaxed except smoothness of $f(x-\theta)$. Consider a first order MIPA

AD-A147 891 BASIC RESEARCH IN ELECTRONIC (JSEP): JOINT SERVICES 3/3
ELECTRONICS PROGRAM..(U) POLYTECHNIC INST OF NEW YORK
BROOKLYN MICROWAVE RESEARCH INST.. A A OLINER
UNCLASSIFIED 30 SEP 84 POLY-MRI-1432-83 AFOSR-TR-84-0978 F/G 9/5 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

$$C_1^*(x) = \sum_{j=1}^{m} (C_{1j}x + C_{0j})I_{A_j}(x) \tag{D.3}$$

where $A_j=(a_{j-1},a_j]m$ $h = 1,2,\ldots, j, m; a_0 = -\infty$, $a_m = +\infty$ and $\int_{a_{j-1}}^{a_j} f(x)dx = \frac{1}{m}$.

The mean-square error E_r^* beweeen $\psi_1^*(x)$ and $-f^1(x)/f(x)$ is

$$E_r^* = \sum_{j=1}^{m} \left[C_{1j}^2 \int_{a_{j-1}}^{a_j} x^2 f(x)dx + 2C_{1j}C_{0j} \int_{a_{j-1}}^{a_j} xf(x)dx \right.$$

$$+ C_{0j}^2 \int_{a_{j-1}}^{a_j} f(x)dx + 2C_{1j} \int_{a_{j-1}}^{a_j} xf^1(x)dx + 2C_{0j} \int_{a_{j-1}}^{a_j} f^1(x)dx$$

$$+ \int_{-\infty}^{\infty} [f^1(x)/f(x)]^2 f(x)dx$$

The scores C_{1j} and C_{0j} which minimize E_r^* are

$$C_{1j} = \frac{\begin{vmatrix} P_{1j} & \mu_{1j} \\ P_{0j} & a_{0j} \end{vmatrix}}{\begin{vmatrix} \mu_{2j} & \mu_{1j} \\ \mu_{1j} & \mu_{0j} \end{vmatrix}} \tag{D.4}$$

$$C_{0j} = \frac{\begin{vmatrix} \mu_{2j} & P_{1j} \\ \mu_{1j} & P_{0j} \end{vmatrix}}{\begin{vmatrix} \mu_{2j} & \mu_{ij} \\ \mu_{1j} & \mu_{0j} \end{vmatrix}} \tag{D.5}$$

where

$$\mu_{2j} = \int_{a_{j-1}}^{a_j} x^2 f(x)dx, \qquad \mu_{1j} = \int_{a_{j-1}}^{a_j} xf(x)dx, \qquad \mu_{0j} = \int_{a_{j-1}}^{a_j} f(x)dx$$

$$P_{1j} = \int_{a_{j-1}}^{a_j} f(x)dx + a_{j-1}f(a_{j-1}) - a_j f(a_j), \qquad P_{0j} = [f(a_{j-1}) - f(a_j)]$$

It can be shown that the scores of Eqs. (D.4) and (D.5) also minimize the variance of the estimator with the corresponding

$$V_{min}(\psi_1^*, f) = \sum_{j=1}^{m} (\tilde{C}_{1j} P_{1j} + \tilde{C}_{0j} P_{0j}) .$$

The pertinent parameters can be estimated continuously using a microprocessor operating in a parallel mode based on the recursive procedures presented in Section B.2.

E. Additional Relevant Progress

Some additional recent accomplishments which are relevant to the content of this proposal are included now.

(1) Radar, sonar, and earthsounding images result from particularly weak signals. In addition, the underlying distributions in the latter class of problems are poorly defined. Faced with these difficulties, numerous investigators in this area concentrated their research efforts on array processors, preferably of the robust type.[34 36] Some new results in this area were presented in a paper.[33] The typical assumptions of gaussianity and stationarity of additive noise were removed. In addition, it was assumed that the noise at the sensor inputs are intercorrelated. To robustize the array processing procedure, the optimum nonlinearities at the correlator's inputs were approximated by simplified nonlinear functions based on parameters which can be easily estimated. The effect of the inter-sensor noise dependence on the asymptotic relative efficiency (ARE) of a multiple-input correlation detection procedure was examined. Easy-to-implement detector realizations were shown to be efficient when compared with the optimum detector configurations.

(2) In last year's report three classes of problems -- object reconstruction algorithms,[18] extraction of moving objects by recursive parallelepiped masks,[7] and edge detection using Latin square masks[6] -- were described, which depended strongly on a good robust recursive vector parameter estimator. A refined version of such an estimator was generated.[22] An improved rate of convergence and guaranteed low variance of estimation and high degree of robustness (insensitivity to variations in underlying noise distributions) result from the use of rank-type preprocessors. The latter class of preprocessors act as efficient gaussianizers of poorly conditioned data.

(3) The results on dependent sampling in sequential partition detectors reported in reference 12 were extended to a more general model of Markov-dependent noise.[14] This paper involves proving certain basic theorems which point to the fact that considerable advantage in detection efficiency of partition detectors is achieved if one uses the Markov-dependent structure of noise. The results open the door to a whole new class of robust detectors which utilize the structure of dependence of the underlying noise to improve the efficiency of detection.

(4) The advent of modern high-speed sampling techniques resulted in additional stress on the problem of signal processing in dependent noise mentioned in the previous paragraph. Since in many systems of practical interest the knowledge of the noise statistics is either inexact or unspecified, some adaptive techniques which are frequently ineffective or too complicated to implement were suggested in the past. However, if one applies the methodology of system identification, the approach is essentially equivalent to the operation of a prewhitening filter. In particular, in a recent paper[17] a procedure for robust identification of the autoregressive (AR) model was introduced. The robustizing process is in the form of a robustized Robbins-Monro stochastic approximation (RMSA) algorithm and is based on the m-interval polynomial approximation method. The resulting algorithm represents a recursive robustized version of the well-known maximum entropy method (MEM) for spectral estimation introduced by Burg[37] or of the popular Widrow least-mean-square (LMS) adaptive filter.[38] Furthermore, the robustized algorithm leads naturally to a robustized Akaike's information criterion (AIC).[39] The simplicity of implementation and flexibility make the application of the new robust identification algorithm particularly attractive in practical applications. The flexibility and robustness of the new procedure were confirmed by extensive Monte Carlo simulations. Some of these methods will be tested on real underwater sounding data obtained from the Navy.

(5) The standard Kalman filter will usually diverge when the dynamic noise is not gaussian. A robust Kalman filter based on m-interval polynomial approximation (MIPA) method for nongaussian noise was formulated.[16] Two situations were studied in detail: (a) the states are gaussian and the observation noise is nongaussian, (b) the states are nongaussian and the observation noise is gaussian. It was shown that the MIPA Kalman filter maintains efficient performance over a broad class of noise distributions and is both computationally attractive and easy to implement. Compared to the min-max Kalman filters suggested by other authors, the new class of filters is more efficient and robust. The theoretical results were verified by extensive Monte Carlo simulations.

(6) A particularly efficient and flexible new approach to robust detection was introduced in reference 15. In the paper, the Huber-Tukey model of mixture distributions is replaced by a more versatile one which leads to solutions which are more appealing from a practical point of view than the robust detectors using the Huber-Tukey noise model in conjunction with the min-max criterion of performance. The new detector consists of two parts -- parametric and nonparametric -- which are switched in or out depending on one parameter of the mixture distribution model. As part of the detector, an efficient running estimator of this unknown parameter is included. When compared to the min-max detector proposed by other researchers, the new detector is more efficient for deterministic and stochastic signals. This divergence in efficiencies is particularly apparent under severe and/or variable noise conditions.

(7) The theory of partition detectors was extended to so-called double-threshold detectors which are particularly useful in radar applications.[41] Quantization loss of the commonly used double-threshold detector is minimized by using a multilevel threshold for the first threshold. The overall performance is improved by optimal selection of

partitions (thresholds) and scores. Expressions for false alarm rates and detection probabilities and upper bounds based on the generalized Chebyshev inequality are presented. In radar applications, the new detector performs better than detectors suggested by other researchers. The new detector has potential for application to communications through dispersive media.

(8) The problem of estimating optimum scores for some robust detectors (classifiers) was considered.[42] The detectors of interest are those in the class of partition detectors. They are robust in the sense that a change in the environment will not degrade their performance significantly. The RMSA algorithm was used to obtain recursive estimators. The difficulty with this approach lies in choosing suitable regression functions which are not unique. Two approaches using direct and indirect methods were conceptualized and shown to be sufficiently flexible to yield efficient estimators of scores for almost all partition detectors of practical interest.

(9) The theory of nonlinear and recursive equalizers, presented by the principal investigator in separate papers in the past, was unified to yield nonlinear recursive equalizers which retain good qualities of both classes: immunity to burst noise, and improved efficiency and flexibility.[43] It was shown that it is not always advantageous to reduce the intersymbol interference to zero if the overall mean-square error is to be minimized in the presence of intersymbol interference and noise. The design procedure involved a proper selection of pole-zero locations to ensure optimum performance with minimum sensitivity to selection of parameters in the equalizer. Specific examples were presented in support of the theory developed in the paper.

(10) An orthogonal-mixing concept for adaptive detection in incompletely specified noise was introduced.[44] This is a natural continuation of the robustness principles introduced by the principal investigator through the years. Inherent robustness of some detectors (classifiers) of the LRO or LMIP class was combined with the high efficiency of parametric detectors to obtain good overall performance. A basic theorem on orthogonally combining detectors was proven, which leads naturally to the design of numerous practical, efficient, and robust detectors.

(11) The theory of MIPA detection was extended to sequential operation of the detector.[47] The detector operates at near-optimum levels for a particular noise environment and is robust by maintaining high efficiency in other than nominal noise environments by adapting its optimum nonlinearity using an m-interval approximation to it. It was shown that the sequential MIPA detectors are asymptotically optimum, increase their transmission rate up to four times as compared to their fixed-sample-size counterparts, and are highly insensitive to variation of noise if compared to detectors based on min-max theory.

(12) A further improvement in efficiency was shown to be possible in operating robust detectors in a sequential mode. The basic structure of the new class of sequential detectors is based on the theory introduced by Cochrane and Kurz.[49] Though the detector is more complicated than the structure introduced in reference 47, its performance may be improved by as much as 3 db in comparison with MIPA detectors,

and the flexibility of design permits inclusion of side constraints in the design procedure.[48]

(13) The concepts of local and global robustness were introduced, which improve our ability to design robust detectors and/or estimators with desirable properties in various regions of operation. Preliminary results on the subject will be presented at the 1983 Symposium on Information Theory.[50]

(14) A recursive receiver structure for receiving Binary Phase Shift Keyed (BPSK) signals over bandlimited, nonlinear satellite channels in the presence of the downlink additive Gaussian noise was derived. The optimum recursive equalizer structure was obtained using the Wiener-Kolmogorov theory. In analyzing the problem, the decision is made on a typical signal in a received sequence taking into account past and future interfering signals, i.e., I.S.I. As an illustrative example of the receiver, a soft limiting channel model was considered in detail. Based on the analysis, an alternative receiver structure which is more suitable for implementation was introduced. The performance of the receiver was evaluated using extensive computer simulation. The new receiver structure yields significant improvement in performance when compared with a single sample sign detector. Preliminary results pertaining to this satellite communication system were presented to the 1983 IEEE Globcom Conference (see reference 52).

(15) A summary of some recent contributions to the theory of sequential partition detectors was given in a recent invited talk and paper (see reference 53). The paper also includes some new contributions, i.e., the use of Lehmann alternatives in optimization and evaluation of robust detectors, and the choice of some optimum truncation techniques for proper operation of practical sequential detectors.

5. REFERENCES

1. K.S. Fu and A. Rosenfeld, "Pattern Recognition and Image Processing," IEEE Trans. Comp., Vol. C-2, pp. 358-368 (December 1976).

2. O. Kempthorne, The Design and Analysis of Experiments, Wiley (1952).

3. W.G. Cochran and G.M. Cox, Experimental Designs, Chapter 1, 2nd Ed., Wiley (1957).

4. R.A. Fisher, The Design of Experiments, Hafner, pp. 17-21 (1960).

5. C.A. Mohwinkel and L. Kurz, "Computer Picture Processing and Enhancement by Localized Operations," Computer Graphics and Image Processing, 5, pp. 401-424 (1976).

6. I. Kadar and L. Kurz, "A Class of Robust Edge Detectors Based on Latin Squares," Pattern Recognition, Vol. 11, pp. 329-339 (1974).

7. I. Kadar and L. Kurz, "A Class of Three-Dimensional Recursive Parallelepiped Masks," Computer Graphics and Image Processing, 11, pp. 262-280 (1979).

8. S. Kariolis and L. Kurz, "Object Detection and Extraction: A Statistical Approach," Proc. 1980 Conf. Syst. and Info. Sci., Princeton, pp. 500-505 (March 1980).

9. L. Kurz, "Nonparametric Detectors Based on Partition Tests," in Nonparametric Methods in Communications: Selected Topics, P. Papantoni-Kazakes and D. Kazakes, eds., Marcel Dekker (1977).

10. P. Kersten and L. Kurz, "Improved Operation of m-Interval Detectors by Optimum Signal Selection," IEEE Trans. on Info. Theory, Vol. IE-24, No. 4 (July 1978).

11. P. Kersten and L. Kurz, "Bivariate m-Interval Classifiers with Application to Edge Detection," Information and Control, Vol. 34, No. 2, pp. 152-168 (June 1977).

12. R.F. Dwyer and L. Kurz, "Sequential Partition Detectors with Dependent Sampling," J. Cybernetics, Vol. 10, pp. 211-232 (1980).

13. R.F. Dwyer and L. Kurz, "Sequential Partition Detectors in Large Signal and Impulsive Noise," Proc. 1980 Conf. Info. Sci. and Syst., Princeton, pp. 507-512 (March 1980).

14. R.F. Dwyer and L. Kurz, "Theory of Partition Detectors for Data Represented by Markov Chains," IEEE Trans. on Info. Theory (in review).

15. L. Kurz and C. Tsai, "Robust Detection in Mixture Noise," IEEE Trans. on Info. Theory (in review).

16. L. Kurz and C. Tsai, "An Adaptive Robustizing Approach to Kalman Filtering," Automatica, April 1983, pp. 279-288.

17. L. Kurz and C. Tsai, "A Robustized Maximum Entropy Approach to System Identification," IFIP Meeting, New York City (September 1981), also Proceedings of the Meeting, Springer Verlag, pp. 276-284.

18. I. Kadar and L. Kurz, "A Robustized Vector Recursive Stabilizer Algorithm for Image Restoration," Information and Control, 44, pp. 320-338 (1980).

19. J. Evans, P. Kersten and L. Kurz, "Robust Recursive Estimation with Applications," Info. Sci., Vol. 11, No. 1, pp. 69-92 (August 1976).

20. P. Kersten and L. Kurz, "Robustized Vector Robbins-Monro Algorithm with Applications to M-Interval Detection," Info. Sci., Vol. 11, No. 2, pp. 121-140 (October 1976).

21. I. Kadar and L. Kurz, "Robustized Scalar Form of Gladyshev's Theorem with Application to Nonlinear Systems," Proc. 1980 Conf. Info. Sci. and Syst., Princeton, pp. 297-302 (March 1980).

22. I. Kadar and L. Kurz, "A Robustized Vector Recursive Algorithm in Estimation and Image Processing," 1981 IEEE Int. Symp. on Info. Theory, Santa Monica, California (February 1981).

23. A.M. Mood and F.A. Graybill, Introduction to the Theory of Statistics, McGraw-Hill (1962).

24. N.J. Johnson and F.C. Leone, Statistical and Experimental Designs in Engineering and Physical Sciences, 2nd Ed., Vol. II, Chapter 14, Wiley (1977).

25. E.L. Lehmann, Testing Statistical Hypotheses, Wiley (1959).

26. G.E.P. Box and S.L. Anderson, "Permutation Theory in the Deviation of Robust Criteria and Study of Departures from Assumptions," J.R. Stat. Soc., Series B, 17, pp. 1-34 (1955).

27. M.N. Woinsky, "Nonparametric Detection Using Spectral Data," IEEE Trans Info. Theory, Vol. IT-18, pp. 110-118 (1972).

28. H.C. Andrews and B.R. Hunt, Digital Image Restoration, Prentice-Hall (1977).

29. G.T. Herman, "Two Direct Methods for Reconstructing Pictures from Their Projections: A Comparative Study," Computer Graphics and Image Processing, Vol. I, No. 2, pp. 124-145 (1972).

30. P.J. Huber, "Robust Estimation of a Location Parameter," Ann. Math. Stat., Vol. 35 (1964).

31. A. Dvoretzky, "On Stochastic Approximation," 3rd Berkeley Symp., Vol. I, pp. 39-55 (1965).

32. J. Sacks, "Asymptotic Distribution of Stochastic Approximation Procedures," Ann. Math. Stat., Vol. 29, pp. 373-405 (1958).

33. L. Kurz and R.A. Mihajlović, "Robustized Weak Random Signal Array Detection," 1981 IEEE International Geoscience and Remote Sensing Symposium (IGARSS '81) Digest, Vol. II, pp. 858-863 (June 1981).

34. J. Capon, "Application of Detection and Estimation Theory to Large Array Seismology," Proc. IEEE, Vol. 58, pp. 760-770 (1970).

35. G.S. Shin and S.A. Kassam, "Multilevel Coincidence Correlators for Random Signal Detection," IEEE Trans on Info. Theory, Vol. 25, pp. 47-53 (1979).

36. L.C. Wood and S. Treitel, "Seismic Signal Processing," Proc. IEEE, Vol. 63, 1975, pp. 649-661 (1975).

37. J.P. Burg, "Maximum Entropy Spectral Analysis," 37th Annual Meeting Soc. Expror. Geophys., Okalhoma City, Oklahoma (1967).

38. B. Widrow, et al., "Adaptive Antenna Systems," Proc. IEEE, Vol. 55, pp. 2143-59 (1967).

39. M. Akaike, "Fitting Autoregressive Models for Prediction," Ann. Inst. Stat. Math., Vol. 21, pp. 243-247 (1969).

40. L. Kurz and C. Tsai, "Robust Detection in Mixture Noise," IEEE Trans. Info. Theory (in review).

41. S.K. Gambhir and L. Kurz, "Double-Threshold Partition Detectors with Radar Applications," IEEE Trans. on Aerospace & Electr. (in review).

42. L. Kurz and B.P. Rabinowitz, "Estimation of Optimum Scores for Partition Detectors Using Stochastic Approximation Techniques", Proc. 16th Annual Conference on Information Sciences and Systems, March 1982, pp. 598-604.

43. N. Huang and L. Kurz, "An Approach to Nonlinear Recursive Equalization in Data Transmission Systems", Proc. 16th Annual Conference on Information Sciences and Systems, March 1982, pp. 145-150.

44. R.A. Mihajlović and L. Kurz, "Orthogonal-Mixing Adaptive Detection in Incompletely Specified Noise", 1982 IEEE International Conference on Communications Record, Vol. I, pp. 1H-4.1 - 1H-4.6, June 1982.

45. E.S.H. Chang and L. Kurz, "Object Detection and Experimental Designs" (in review).

46. E.S.H. Chang and L. Kurz, "Trajectory Detection and Experimental Designs," Computer Vision, Image Processing and Computer Graphics (galley proofs are available).

47. E. Voudouri and L. Kurz, "Sequential Robust m-Interval Polynomial Approximation (MIPA) Partition Detectors," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, April 1983, pp. 611-614.

48. E. Voudouri and L. Kurz, "Generalized Sequential Linear Rank and Partition Tests," Proc. IEEE Int. Conf. on Communications, June 1983, pp. 1597-1603.

49. J. Cochrane and L. Kurz, "Generalized Sign-Test Classifiers Within an Unified Framework of M-Sample Nonparametric Tests," Proc. 1972 Princeton Conference on Info. Sciences and Systems, March 1972, pp. 479-483.

50. R.A. Mihajlović and L. Kurz, "On Local Robustness of Some Statistical Tests," presented at the IEEE Symp. on Info. Theory, Sept. 1983.

51. C.-L. Lin, "A New Decision Feedback Error Control System with Application in Image Processing," Ph.D. Dissertation, PINY, June 1983.

52. A.F. Elrefaie and L. Kurz, "A Recursive Approach to Detection of BPSK in the Presence of Nonlinearity and Gaussian Noise," Proc. IEEE Global Telecommunication Conference, pp. 1039-1044 (Dec. 1983).

53. L. Kurz, "Sequential Partition Detection," presented by invitation to the 1984 IEEE Southwest Symposium on Systems and also to appear in the Proceedings.

END

FILMED

1-85

DTIC